



Facultad de Ciencias Bioquímicas y Farmacéuticas - Universidad Nacional  
de Rosario

Tesis de Doctorado

**Estudios sobre la regulación de la expresión génica por  
microARNs en plantas mediante estrategias  
bioinformáticas**

Presentada por: Uciel Pablo Chorostecki

Rosario, Argentina

2016

# **Estudios sobre la regulación de la expresión génica por microARNs en plantas mediante estrategias bioinformáticas**

Uciel Pablo Chorostecki

Licenciado en Ciencias de la Computación

Universidad Nacional de Rosario

Esta Tesis es presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Biológicas, de la Universidad de Rosario y no ha sido presentada previamente para la obtención de otro título en esta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), dependiente de la Facultad de Cs. Bioquímicas y Farmacéuticas, durante el período comprendido entre Abril de 2012 y Marzo de 2016, bajo la dirección del Dr. Javier Palatnik.

Parte de los resultados de esta tesis fueron publicados en los siguientes artículos:

- Mateos JL, Bologna NG, Chorostecki U, Palatnik JF. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. **Curr Biol**. Jan 12;20(1):49-54. doi: 10.1016/j.cub.2009.10.072.
- Chorostecki U, Crosa VA, Lodeyro AF, Bologna NG, Martin AP, Carrillo N, Schommer C, Palatnik JF. (2012). Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Res*. Oct;40(18):8893-904. doi: 10.1093/nar/gks625.
- Bologna NG, Schapire AL, Zhai J, Chorostecki U, Boisbouvier J, Meyers BC, Palatnik JF. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. **Genome Res**. Oct;23(10):1675-89. doi: 10.1101/gr.153387.112.
- Chorostecki U, Palatnik JF. (2014). comTAR: a web tool for the prediction and characterization of conserved microRNA targets in plants. **Bioinformatics**. Jul 15;30(14):2066-7. doi: 10.1093/bioinformatics/btu147.

## **Agradecimientos**

Agradezco enormemente haber tenido la posibilidad de realizar esta increíble experiencia. Pero el agradecimiento no es sólo por las ayudas a nivel académico sino también para todos los que me acompañaron de una u otra manera durante esta etapa de mi vida.

A mi director, Javier Palatnik por darme la oportunidad de formar parte de su grupo. Su confianza y apoyo en mi trabajo junto con su capacidad para guiar mis ideas ha sido un aporte invaluable, no sólo en el desarrollo de esta Tesis, sino también en mi formación como investigador.

A mis compañeros de grupo. A Belén y Flor por su amistad y por sus aportes al desarrollo de este trabajo. A Carla, Rama, Mati, Ara, Cami, Ana y Anto por las discusiones durante los seminarios y los almuerzos compartidos. A los que ya no forman parte del grupo, pero que fueron muy importante durante mi Tesis y mi formación. En especial a Juan, Nico y Juli por las incontables ayudas, por la amistad que formamos y por las comunicaciones a la distancia que aún mantenemos. A Edgard, Vale, Ana y Sil por sus ayudas en el principio de este camino. A Seba, compañeros en el IBR, por los almuerzos y charlas compartidas. Al grupo Libreciencia, por compartir su pasión por la ciencia.

A toda mi familia por su apoyo. En particular a mis viejos por darme la confianza y libertad de dejarme elegir mi propio camino. A mis hermanos por bancarme en el día a día y por sentir mi felicidad como propias. A mis abuelas que siempre me acompañaron y confiaron en mí. A mi otra familia, la política. En especial a mis cuñados, por demostrar siempre interés por lo que hago y por estar pendientes. A mis primos políticos, por tantos domingos juntos y tantas cenas compartidas. También a mis sobrinos que con su alegría, sus planteos disparatados, sus frases elocuentes me alegran a diario, estando cerca o a la distancia.

A mis amigos de siempre, los del Poli, por dejarme compartir estos años con ustedes, por las charlas, por los asados, por los viajes, por su tiempo, por entenderme, por dejarme



ser parte. A “La Balón”, por todos los sábados de fútbol, los costillares, las enriquecedoras sobremesas y por hacerme sentir uno de los suyos.

A los doctores Lucas Daurelio y Esteban Serra por su tiempo y dedicación como tutores, por sus ayudas y consejos brindados. Al IBR en general y al Dr. Alejandro Vila en particular, por su generosidad para compartir su experiencia y pasión por lo que hacemos.

A el Dr. Cedric Notredame, por abrirme la puerta de su laboratorio en el CRG, Barcelona y por compartir su amplio conocimiento y a la IUBMB por la beca otorgada para poder realizar la estancia.

A CONICET por el apoyo económico.

Finalmente a Caro, por su incondicional apoyo, por su paciencia, por acompañarme donde sea. Por los momentos ya compartidos y porque recién empezamos a contar nuestra historia. Por lo que se nos viene, porque sin dudas lo mejor está(n) por llegar.

# Índice general

<b>Índice de figuras</b>	<b>X</b>
<b>Índice de tablas</b>	<b>XII</b>
<b>Nomenclatura</b>	<b>XIII</b>
<b>Introducción</b>	<b>1</b>
1.1. Silenciamiento génico y ARN pequeños en plantas. . . . .	1
1.1.1. Variedades de ARN pequeños en plantas. . . . .	1
1.2. miARNs en plantas. . . . .	4
1.3. Biogénesis de miARNs. . . . .	4
1.3.1. Procesamiento de miARNs en animales. . . . .	4
1.3.2. Procesamiento de miARNs en plantas. . . . .	5
1.4. Estructuras secundarias de precursores de miARNs. . . . .	7
1.5. Regulación de la expresión génica por miARNs. . . . .	10
1.5.1. Regulación por corte del ARN blanco. . . . .	10
1.5.2. Regulación de la traducción por miARNs. . . . .	11
1.5.3. Generación de ta-siRNAs y siRNAs secundarios. . . . .	11
1.6. Procesos biológicos regulados por miARNs en plantas. . . . .	12
1.7. Predicción de genes blanco de miARNs. . . . .	13
1.7.1. Interacción de miARN-ARNm blanco. . . . .	13
1.7.2. Métodos para predicción de genes blanco de miARNs en animales. . . . .	14
1.7.3. Métodos para predicción de genes blanco de miARNs en plantas. . . . .	15
1.7.4. Identificación de genes blanco regulados por miARNs mediante técnicas de secuenciación de alto rendimiento. . . . .	16

<b>Objetivos</b>	<b>18</b>
2.1. Objetivo general. . . . .	18
2.2. Objetivos específicos. . . . .	19
<b>Materiales y Métodos</b>	<b>20</b>
3.1. Métodos utilizados para la predicción de genes regulados por miARNs en plantas. . . . .	20
3.1.1. MiARN consensos. . . . .	20
3.1.2. Predicción de genes regulados por miARNs. . . . .	20
3.2. ComTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas. . . . .	22
3.2.1. MiARN y transcriptos. . . . .	22
3.2.2. Búsqueda de genes blanco. . . . .	23
3.2.3. Herramienta web y almacenamiento de datos. . . . .	23
3.3. Análisis bioinformático de los datos obtenidos de la secuenciación por la técnica de SPARE. . . . .	24
3.4. Desarrollo de una herramienta intuitiva para el análisis y visualización de los datos obtenidos por la técnica de SPARE. . . . .	24
3.5. Análisis estructural de precursores de miARNs en plantas. . . . .	25
3.6. Estrategia bioinformática para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento. . . . .	26
3.7. Código fuente de los scripts y herramientas desarrolladas. . . . .	29
3.8. Anexo: Técnicas de Análisis Molecular y bioquímico relacionadas al trabajo de Tesis. . . . .	29
3.9. Obtención de Arabidopsis transformadas. . . . .	30
3.9.1. Ecotipos utilizados y condiciones de crecimiento. . . . .	30
3.9.2. Transformación de Arabidopsis. . . . .	30
3.9.3. Recolección y esterilización de semillas. . . . .	31
3.9.4. Selección de transformantes. . . . .	31
3.9.5. Transformación nuclear de tabaco. . . . .	31
3.10. Cuantificación del nivel de expresión génica. . . . .	32
3.10.1. Extracción de ARN. . . . .	32
3.10.2. Cuantificación y chequeo de la integridad del ARN purificado. . . . .	33
3.10.3. Tratamiento del ARN preparado con ADNasa. . . . .	33

3.10.4. Retrotranscripción (RT). . . . .	33
3.10.5. Reacción en cadena de la polimerasa en tiempo real. . . . .	34
3.11. Mapeo del sitio de corte. . . . .	35
3.12. Protocolo de SPARE. . . . .	37
<b>Capítulo 1</b>	<b>38</b>
4.1. Introducción. . . . .	38
4.2. Resultados y Discusión. . . . .	40
4.2.1. Selección de secuencias de miARNs presentes en distintas especies. . . . .	40
4.2.2. Selección de la biblioteca de transcritos de plantas. . . . .	41
4.2.3. Primera búsqueda general de potenciales genes blancos. . . . .	42
4.2.4. Parámetros empíricos y de conservación evolutiva pueden actuar de manera sinérgica para identificar genes blancos regulados por miARNs. . . . .	45
4.2.5. Identificación de nuevos genes blancos en <i>A. thaliana</i> por conservación de la secuencia del gen blanco. . . . .	50
4.2.6. Identificación de nuevos genes blancos permitiendo interacciones G-U. . . . .	53
4.2.7. Identificación de genes blancos específicos de <i>Solanaceae</i> . . . . .	55
4.2.8. comTAR: una herramienta para la predicción de genes blancos regulados por miARNs en plantas. . . . .	57
4.3. Conclusiones. . . . .	59
<b>Capítulo 2</b>	<b>61</b>
5.1. Introducción. . . . .	61
5.1.1. Construcción de bibliotecas de SPARE para estudios genómicos de biogénesis de miARNs en plantas. . . . .	62
5.1.2. Visualización de precursores que se procesan desde la base. . . . .	63
5.1.3. Visualización de precursores que se procesan desde el loop. . . . .	65
5.2. Resultados y Discusión. . . . .	66
5.2.1. Desarrollo de herramienta para el análisis de intermediarios de procesamiento en <i>A. thaliana</i> . . . . .	66
5.2.2. Análisis estructural de precursores de miARNs en plantas. . . . .	71
5.3. Conclusiones. . . . .	73
<b>Capítulo 3</b>	<b>75</b>
6.1. Introducción. . . . .	75

6.2. Resultados y Discusión. . . . .	76
6.2.1. Estrategia bioinformática para el estudio de la evolución y biogénesis de miARNs en plantas. . . . .	76
6.2.2. Visualización por Circos de precursores que se procesan desde la base. . . . .	82
6.2.3. Visualización por Circos de precursores que se procesan desde la base en forma secuencial. . . . .	84
6.2.4. Visualización por Circos de precursores que se procesan de el loop cortos. . . . .	86
6.2.5. Visualización por Circos de precursores que se procesan de arriba secuenciales. . . . .	86
6.2.6. El tamaño del loop es muy variable en precursores que se procesan desde abajo. . . . .	89
6.2.7. Procesamiento mixto de miembros de la familia del miR170/miR171. . . . .	92
6.2.8. Visualización por Circos de precursores con mutaciones puntuales que afectan el procesamiento de miARNs en plantas. . . . .	94
6.2.9. Visualización por Circos de precursores de plantas conservados en monocotiledóneas. . . . .	97
6.2.10. Estrategia bioinformática para estudiar precursores de miARNs en animales. . . . .	99
6.3. Conclusiones. . . . .	103
<b>Conclusiones finales</b>	<b>104</b>
8.1. Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco. . . . .	104
8.2. Desarrollo de la herramienta comTAR. . . . .	105
8.3. Determinantes mecánicos del procesamiento de miARNs en animales. . . . .	105
8.4. Desarrollo de una herramienta para el análisis de bibliotecas de SPARE incluyendo una interfaz gráfica. . . . .	106
8.5. Análisis de las estructuras de los precursores y su evolución. Visualización de información compleja por adaptación de una herramienta Circos. . . . .	107
<b>Bibliografía.</b>	<b>108</b>
<b>Anexo</b>	<b>120</b>

# Índice de figuras

1.1. Familias de miARNs conservados en plantas . . . . .	3
1.2. Biogénesis de miARNs en animales . . . . .	6
1.3. Biogénesis y actividad de los miARNs en plantas . . . . .	8
1.4. Tamaños de precursores de miARNs . . . . .	9
1.5. Estructuras de precursores de miARNs . . . . .	10
1.6. Interacción de miARN-ARNm blanco . . . . .	13
3.1. comTAR. Diagrama de flujo . . . . .	23
3.2. Análisis informático utilizado para procesar los datos de SPARE . . . . .	25
3.3. Especies almacenadas en Phytozome . . . . .	27
3.4. Diagrama de la estrategia utilizada para la identificación de los productos de corte de miARNs . . . . .	36
3.5. Esquema general para la construcción de bibliotecas de SPARE . . . . .	37
4.1. Conservación y divergencia de miARNs en distintas especies . . . . .	39
4.2. Estrategia . . . . .	44
4.3. Conservación de potenciales genes blancos en distintas especies . . . . .	46
4.4. Selección de genes blancos por conservación evolutiva de la secuencia . . . . .	48
4.5. Nuevos genes blancos validados en <i>A. thaliana</i> . . . . .	52
4.6. Identificación de un nuevo gen blanco de miARN . . . . .	54
4.7. Identificación de genes blancos de miARN, específicos de <i>Solanaceae</i> . . . . .	56
4.8. Resultados del comTAR para el miR398 . . . . .	58
4.9. Identificación de genes regulados por un miARN en particular usando las herramienta comTAR . . . . .	58
5.1. Variación en el tamaño de precursores de miARNs de plantas . . . . .	62
5.2. Esquema del procedimiento para analizar los datos de SPARE . . . . .	63
5.3. Técnica de SPARE . . . . .	64

5.4. Identificación y caracterización de precursores de miARNs procesados de base a loop . . . . .	64
5.5. Identificación y caracterización de precursores de miARNs procesados de loop a base . . . . .	65
5.6. Captura de pantalla de la herramienta de SPARE para el miR165a . . . . .	68
5.7. Captura de pantalla de la herramienta de SPARE para el miR159b . . . . .	70
5.8. Estructura secundaria de precursores de base a loop . . . . .	72
5.9. Estructura secundaria de precursores de loop a base . . . . .	72
5.10. Mecanismos de procesamiento . . . . .	73
6.1. Estructura secundaria del miR172a en distintas especies . . . . .	78
6.2. Alineamiento del precursor del miR172a. . . . .	79
6.3. Circos del miR172a . . . . .	81
6.4. Circos del miR172a . . . . .	83
6.5. Circos de precursores con mecanismos de procesamientos secuenciales de base a loop . . . . .	85
6.6. Circos de precursores con mecanismos de procesamientos cortos de loop a base . . . . .	87
6.7. Circos de precursores con mecanismos de procesamientos secuenciales de loop a base . . . . .	88
6.8. Variación del loop terminal en precursores miR160 . . . . .	90
6.9. Variación del loop terminal en precursores miR172a . . . . .	91
6.10. Patrón de conservación de la familia con procesamiento mixto del miR170/miR171 . . . . .	93
6.11. Dúplex del miR164a/miR164a* y estructura secundaria del precursor . . . . .	94
6.12. Circos del miR164a . . . . .	95
6.13. Patrón de conservación del precursor del miR394b . . . . .	96
6.14. Circos del miR172a . . . . .	98
6.15. Patrón de conservación del precursor del precursor del miARN hsa-let-7a-1 . . . . .	100
6.16. Circos de animales vs plantas . . . . .	102
9.1. Circos de dicotiledóneas (en más de 20 especies). 1/5 . . . . .	123
9.2. Circos de dicotiledóneas (en más de 20 especies). 2/5 . . . . .	124
9.3. Circos de dicotiledóneas (en más de 20 especies). 3/5 . . . . .	125
9.4. Circos de dicotiledóneas (en más de 20 especies). 4/5 . . . . .	126
9.5. Circos de dicotiledóneas (en más de 20 especies). 5/5 . . . . .	127

# Índice de tablas

3.1. Programa de termociclado . . . . .	35
3.2. Oligos utilizados para la RT-qPCR . . . . .	35
3.3. Oligonucleótidos utilizados para 5' RACE . . . . .	37
4.1. miARNs y sus genes blancos en plantas . . . . .	41
4.2. Especies y base de datos utilizadas para la búsquedas de genes blancos de miARNs conservados . . . . .	43
4.3. Detección de genes blancos de miARNs utilizando diferentes filtros . . . . .	49
5.1. Bibliotecas de SPARE . . . . .	67
6.1. Especies de <i>Metazoa</i> utilizadas . . . . .	101
9.1. Genes blanco validados de familias miARNs conservados en plantas . . . . .	128
9.2. Mecanismos de procesamiento de miARNs conservados en <i>A. thaliana</i> . . . . .	129



# Nomenclatura

## Abreviaturas y símbolos

A Adenina

ADN Ácido desoxirribonucleico

ADNasa ADN Nucleasa

ADNc ADN complementario

ARN Ácido ribonucleico

ARNsh ARN doble hebra

ARNsh ARN simple hebra

ATP Adenosín trifosfato

C Citosina

Col-0 Ecotipo Columbia-0 de *Arabidopsis thaliana*

DCL Ribonucleasa III tipo Dicer

G Guanina

HEN1 Hua enhancer 1

HYL1 Hyponastic leaves 1

LB Medio de cultivo para bacterias Luria-Bertani

miARN MicroARN

MS Medio de cultivo para plantas Murashige-Skoog

nat-miARN miARNs naturales

nat-siARN siARNs naturales

nt Nucleótidos

PCR Reacción en cadena de la polimerasa

pre-miARN Precursor de miARN

RISC Complejo de silenciamiento inducido por ARN

RT-qPCR Transcripción reversa seguida de PCR en tiempo real cuantitativa

RT Retrotranscripción

SE Serrate

siARN ARN pequeño interferente

ta-siARN trans-acting siARNs

T Timina

U Uracilo

### **Lenguaje Especial**

bulge

contigs

dataset

loop

mismatch

scramble

script

software

## Resumen

Los microARNs (o miARNs) son ARN no codificantes que regulan la expresión génica en animales y plantas, y están implicados en procesos biológicos muy variables, como el desarrollo, la diferenciación y el metabolismo. Con un largo de aproximadamente 21 nucleótidos, los miARNs reconocen secuencias parcialmente complementarias en los ARNm blanco, provocando su corte o arresto de la traducción. Los miARNs han saltado rápidamente a la primera plana del interés de la comunidad científica como un nuevo nivel en el control de la expresión génica en eucariotas. Estudios recientes han puesto de manifiesto que los miARNs están implicados en distintas patologías de seres humanos. Los cálculos actuales consideran que cerca del 40 % de los genes de humanos se encuentran regulados por miARNs.

Está generalmente aceptado que los miARNs en plantas tienen una extensiva complementariedad con sus genes blanco y su predicción por lo general se basa en el uso de parámetros empíricos deducidos de interacciones conocidas y validadas experimentalmente. En este trabajo, primero desarrollamos una estrategia para la identificación de genes blanco regulados por miARNs en plantas, basado en la conservación evolutiva del par miARN-gen blanco. Además, pudimos encontrar genes blanco específicos de Solanaceae y demostrar que la estrategia se puede utilizar para la búsqueda de genes blanco pertenecientes a un grupo determinado de especies. Esta estrategia fue usada para predecir nuevas interacciones no conocidas anteriormente en *Arabidopsis thaliana*, que luego fueron validados experimentalmente. Algunos de los nuevos genes identificados podrían participar de las mismas vías que genes blanco conocidos anteriormente, sugiriendo que algunos miARNs pueden controlar diferentes aspectos de un proceso biológico.

A partir de estos resultados, desarrollamos una herramienta bioinformática disponible en un servidor de acceso público para identificar genes blanco de miARNs basada principalmente en la conservación durante la evolución de la interacción del par miARN-gen blanco en distintas especies. La herramienta también brinda una descripción de varios parámetros de

las interacciones miARN-gen blanco en distintas especies, que puede ser útil para mejorar los sistemas de predicción y describir la co-evolución de los miARNs y los genes blanco.

La biogénesis de los miARNs es un proceso clave porque determina la secuencia exacta de nucleótidos del ARN pequeño funcional, que luego determina los genes a ser regulados. Los precursores de miARNs de plantas son muy variables en forma y tamaño en comparación con los precursores de miARNs de animales. Y por lo tanto, poco se sabía sobre el reconocimiento de los precursores por la maquinaria de procesamiento. En la segunda parte de este trabajo presentamos una estrategia para estudiar la conservación de los precursores de miARNs de plantas en distintas especies identificando regiones y dominios presentes en distintas especies. A partir de estos resultados, desarrollamos una herramienta de visualización para poder analizar fácilmente la conservación de la estructura primaria y secundaria de los precursores de miARNs. El análisis de las mismas reveló que existe una marcada correlación entre la conservación de los precursores y el mecanismo de procesamiento. Los resultados muestran patrones de conservación en los dominios necesarios para el procesamiento, y permiten deducir un modelo general del reconocimiento del ARN durante la biogénesis de miARNs.

# Introducción.

## 1.1. Silenciamiento génico y ARN pequeños en plantas.

El control de la expresión génica es un proceso vital para los organismos unicelulares. Este proceso es utilizado por las células para ajustar la expresión génica durante el desarrollo o frente a los cambios en su ambiente. En plantas, la generación de los numerosos tipos celulares diferentes que forman en conjunto un organismo multicelular depende de que los genes se activen en las células adecuadas y en los momentos precisos del desarrollo. Muchos de los mecanismos regulatorios operan a nivel genético generando un control a nivel transcripcional y post-transcripcional.

En los últimos años se ha descubierto que pequeñas moléculas de 20 a 25 nucleótidos de longitud, son reguladores críticos de la expresión génica en eucariotas. En plantas existen distintos tipos de ARN pequeños que, por diferencias en su biogénesis y modos de acción, han sido clasificados en distintas clases, aunque pueden compartir algunos elementos comunes.

### 1.1.1. Variedades de ARN pequeños en plantas.

Los ARN pequeños provienen del procesamiento de regiones helicoidales de precursores de ARN. Los mismos se pueden clasificar en dos grupos principales. Los derivados de precursores de doble hebra (ARNdh) formados por hibridación intermolecular de dos cadenas de ARN complementarias y los derivados de precursores de RNA de una sola hebra (ARNsh) que forman una estructura de tallo-burbuja.

Los ARN pequeños más abundantes, derivados de precursores de ARNdh, son conocidos como siARNs (del inglés small interfering RNAs). Éstos pueden ser de origen endógeno o exógeno. A su vez, los siARNs se pueden clasificar en dos grupos, aquellos que producen un silenciamiento a nivel transcripcional (TGS, del inglés: Transcriptional Gene Silencing)

[31, 148] y aquellos que lo producen a nivel post-transcripcional (PTGS, del inglés: Post Transcriptional Gene Silencing)[60].

El primer grupo se origina a partir de ciertas regiones genómicas como transposones u otras secuencias repetitivas que son transcritos por la ARN polimerasa IV. Estos ARNsh se convierten en ARNdh por acción de una enzima ARN polimerasa dependiente de ARN llamada RDR2 (del inglés RNA-DEPENDENT RNA POLYMERASE2). A continuación los ARNdh son procesados por una ribonucleasa llamada DCL3 (del inglés DICER-LIKE3) que corta el ARNdh para liberar siARNs de 24 nt de longitud. El segundo grupo, los PTGS son originados a partir de ARNs de virus o ARNm proveniente de un transgén. Estos ARNs se convierten en ARNdh por acción de RDR6 y el ARNdh resultante es procesado por DCL2 o DCL4 para finalmente liberar siARN de 21-22 nt que serán encargados del silenciamiento post-transcripcional [156]. Los roles biológicos del silenciamiento mediado por siARNs incluyen resistencia contra virus [147], protección del genoma contra elementos móviles de ADN y roles regulatorios sobre genes endógenos [19, 31, 148].

Existen otros ARNdh menos abundantes que han sido descubiertos posteriormente. Este grupo está constituido por un lado por los ta-siARN (del inglés, trans-acting siRNAs) y por el otro lado los nat-siARNs y nat-miARNs (del inglés, natural antisense siRNAs y miRNAs) [31, 113, 142, 148].

Los genes que darán lugar a ta-siARNs son transcritos para dar lugar a un ARN no codificante denominados TAS. Este transcripto luego es identificado y cortado por acción de un miARN que posee una secuencia complementaria al mismo [5]. Los fragmentos de corte de un TAS se estabilizan y por la acción de RDR6, y se convierten en un ARNdh que es procesado por DCL4, generando ARNs pequeños de 21 nucleótidos. Finalmente, algunos ta-siARNs se incorporan a un complejo RISC y guían el corte de los transcriptos de otros genes de manera similar a la acción de un miARN [5, 51, 154, 163].

En contraste con los otros tipos de siARNs, que dependen de un RDR para sintetizar el precursor ARNdh, se cree que los precursores de ARNdh de NAT-siARNs (siARNs asociados a ARNs antisentido naturales, del inglés “Natural Antisens Transcripts”) [25, 69]. En este caso, dos ARNs independientes pueden ser complementarios debido a que fueron transcritos por hebras opuestas en el mismo locus, y se denominan cis-NAT-siARNs. El ARNdh formado es procesado por DCL1 o DCL2 liberando siARNs primarios que inducen el corte sobre el ARNm expresado constitutivamente. Posteriormente, una RDR sintetiza la hebra complementaria del ARNm cortado. Esto da lugar nuevamente a un ARNdh el cual es

procesado por DCL1 para producir una sucesión de siARN ubicados en fase de 21 nt cada uno denominados nat-siARN secundarios [25, 69].

Como contraste con los siARNs descritos anteriormente, que son procesados por ARNdh, existe la posibilidad de que un ARN se pliegue sobre si mismo, formando un tallo y una burbuja (en inglés, “stem loop”). La clase más abundante de ARN pequeños derivados de precursores de ARNsh (ARN simple hebra) son los miARNs. Tienen entre 20 y 22 nt de longitud y son originados de largos precursores con extensa estructura secundaria en forma tallo-burbuja. Estos precursores luego son procesados por acción de DCL1, liberando el ARN maduro. Actúan como reguladores negativos de la expresión génica reprimiendo la expresión de genes endógenos principalmente a nivel post-transcripcional. Esta regulación se puede dar a través de la inhibición de la traducción o del corte de los ARN mensajeros. Tanto en animales como en plantas, existen un gran número de miARNs conservados en distintas especies. En animales se han encontrado miARNs conservados en especies muy distantes como gusanos hasta humanos [110]. Y en plantas sucede lo mismo, donde existen miARNs conservados desde musgos hasta dicotiledóneas [13, 14, 38] (Figura 1.1). No se han encontrado miARNs conservados en animales y en plantas, lo que sugiere que han aparecido en forma independientes durante la evolución [38]. Si bien los miARNs en plantas y en animales comparten componentes básicos, su biogénesis y su modo de acción difiere entre ellos.

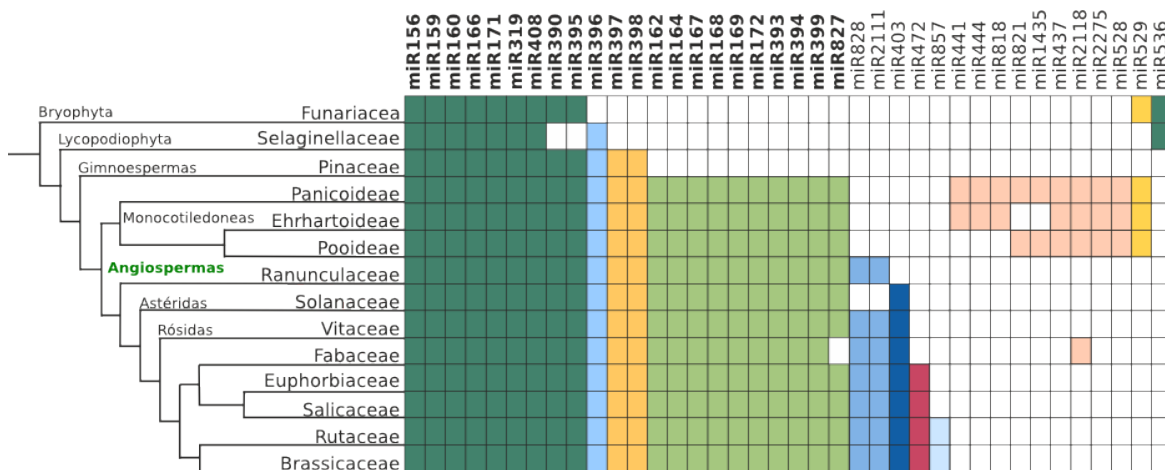


Figura 1.1: **Familias de miARNs conservados en plantas.** Se muestran las familias de miARNs (columnas) y para cada una de ella se muestra en qué especies están conservadas (filas). En diferentes colores se muestra el rango de especies en el que se encuentran conservadas, para distintos grupos de familias. En negrita se marcan las familias conservadas en angiospermas. MiR319 y miR159 que codifican para miARNs similares, fueron considerados como familias diferentes ya que regulan a genes blanco distintos [108]. Adaptado de [38].

## 1.2. miARNs en plantas.

Los miARNs son generados a partir de loci endógenos, tanto en animales como en plantas. Estos ARN pequeños controlan una gran variedad de procesos biológicos, como el desarrollo, la diferenciación y proliferación celular, y respuesta a estrés [12, 30, 38, 128, 148]

Hasta hoy, en *A. thaliana* se han identificado más de 300 [72] miARNs. Se han utilizado distintos enfoques para identificar los miARNs: el clonado directo de ARN pequeños, secuenciación de alto rendimiento, estudios genéticos y predicciones bioinformáticas [38], siendo esta última la más común para la mayoría de las especies.

Los miARNs en plantas están codificados por familias de genes de 1 a 32 miembros que dan lugar a miARNs maduros idénticos o muy similares. Hasta el momento han sido definidas unas 42 familias de miARNs en plantas, las que regulan una amplia variedad de procesos biológicos. Doce de dichas familias tienen como blanco ARN mensajeros que codifican factores de transcripción involucrados en el desarrollo, mientras que otras están relacionadas con rutas de respuesta a señales ambientales y hormonales, entre otros, estando la mayoría de ellas conservadas entre mono y dicotiledóneas [67]. Muchos de estos pequeños ARNs han aparecido recientemente en la evolución y por lo tanto aparecen en un número pequeño de especies [12, 14]. Además, no está claro si estos últimos tienen algún rol biológico [14, 38].

Sin embargo, existen 22 familias de miARNs que están altamente conservadas en las plantas, estando presentes en angiospermas, gimnospermas y algunas de ellas aún en plantas basales como los musgos [10, 12, 165] (Figura 1.1). Muchas de estas familias de miARNs conservados cumplen funciones esenciales para la biología de las plantas [67].

## 1.3. Biogénesis de miARNs.

### 1.3.1. Procesamiento de miARNs en animales.

En animales, el procesamiento de los precursores es realizado en el núcleo y el citoplasma. Luego de la transcripción, el complejo formado por la ribonucleasa de tipo III, DROSHA, y la proteína DiGeorge syndrome critical region gene 8 (DGCR8) produce un primer corte sobre el pri-miARN para liberar una estructura de tallo y burbuja denominado pre-miARN (Figura 1.2) [43, 56, 61, 164].

A continuación, el pre-miARN es traslocado desde el núcleo hacia el citoplasma por la proteína de transporte nuclear EXPORTIN-5 (EXP5) [88, 162]. En el citoplasma DICER



conjuntamente con TRPB, otra proteína de unión a ARNdh, produce un segundo corte sobre el precursor para que liberar el dúplex miARN/miARN\*. Este segundo corte ocurre a ~21 nt de distancia del sitio donde se había producido el primer corte por DROSHA. Finalmente, el miARN maduro es incorporado al complejo RISC [70] (Figura 1.2).

Además de esta ruta canónica de procesamiento, recientemente se han descrito vías alternativas. Para algunos precursores, el requerimiento de DROSHA puede ser suplantado por la maquinaria de corte y empalme (splicing) [105, 122]. En el caso de otros, es necesaria la interacción de proteínas específicas con el loop terminal [58, 141]. Por último, existen casos en donde, el procesamiento de ciertos precursores puede estar específicamente inhibido en ciertos tejidos o determinadas condiciones [64, 104, 146].

### 1.3.2. Procesamiento de miARNs en plantas.

Los miARNs se diferencian de otros ARNs pequeños por su particular biogénesis que implica su escisión de un precursor con extensa estructura secundaria localizado en un largo transcripto primario (Figura 1.3). En general, la biogénesis de estos ARN pequeños comienza con la transcripción por la ARN polimerasa II [154] a partir de unidades transcripcionales propias distribuidas en el genoma [116]. Los transcriptos primarios, llamados pri-miARNs, pueden tener varias kilobases de longitud y sufrir modificaciones post-transcripcionales como ser splicing, capping y poliadenilación. Estos transcriptos contienen precursores para miARNs con extensa estructura secundaria en forma de tallo-burbuja (stem-loop) [67] (Figura 1.3). Donde el miARN maduro se ubica en uno de los brazos del tallo apareado con una secuencia parcialmente complementaria denominada miARN\*.

En animales, el procesamiento comienza en el núcleo por DROSHA y finaliza en el citoplasma por la acción de DICER. En plantas no existe un homólogo de DROSHA. Los precursores son procesados completamente en el núcleo a través de la acción de una ribonucleasa llamada DCL1 [116, 125] (del inglés DICER LIKE 1) en asociación con el cofactor proteico de unión a ARN de doble hebra HYL1 [63] (del inglés HYPONASTIC LEAVES 1) y la proteína SERRATE [86] (Figura 1.3).

Al parecer es la estructura secundaria por sobre la secuencia primaria del precursor la más importante en la determinación del correcto procesamiento del mismo [22]. El producto generado a partir de los cortes llevados a cabo por DCL1, es un dúplex miARN-miARN\* que luego continúa siendo procesado por otros componentes enzimáticos hasta dar lugar

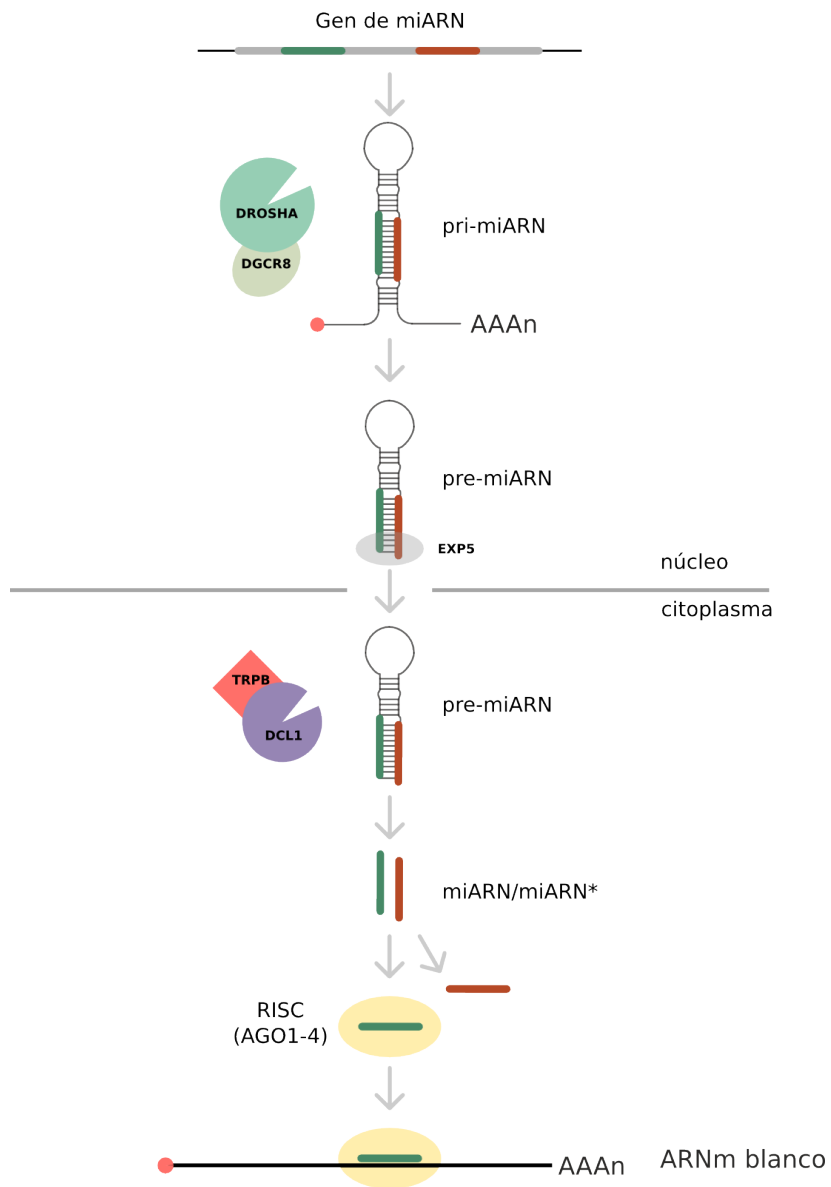


Figura 1.2: **Mecanismo de procesamiento de miARNs en animales.**

al miARN maduro de 21 nt. El miARN\* es originado por el mismo precursor y es casi complementario al miARN.

El paso final de la biogénesis de los miARN es la incorporación asimétrica, a partir del dúplex miARN-miARN\*, del miARN maduro dentro de un complejo de silenciamiento, mientras que el miARN\* en su mayoría se degrada. Este complejo se denomina RISC (del inglés RNAi Silencing Complex). El componente central de todos los complejos de silenciamiento es un miembro de la familia de proteínas ARGONAUTA (AGO). En *Arabidopsis* existen distintas proteínas AGO que participan en diferentes procesos biológicos [143] y la incorporación de los ARN pequeños en los distintos complejos depende de la identidad del nucleótido del extremo 5' y de la vía de biogénesis [96, 99, 138] (Figura 1.3).

En la mayoría de los miARNs el nucleótido extremo 5' es una U y en general la principal efectora de la actividad es AGO1 [96, 144, 148]. Complejos RISC similares se encuentran presentes en células animales. Más recientemente han sido identificadas proteínas adicionales que regularían la actividad de la maquinaria de procesamiento [22].

En animales, los miARNs reconocen principalmente sitios blancos ubicados en la región 3' no codificante de ARN mensajeros blanco inhibiendo su traducción. En cambio en plantas, es más común que los miARNs se unan a secuencias complementarias que se localizan dentro los ARNm blanco en la región codificante señalándolos para su degradación [67]. En cualquier caso, es el miARN el que proporciona la especificidad contra las moléculas de ARN blanco [18].

#### **1.4. Estructuras secundarias de precursores de miARNs.**

La biogénesis de miARNs involucra el adecuado procesamiento del precursor para liberar el miARN maduro. Para que los miARNs puedan actuar de manera correcta en la regulación de la expresión génica, es necesario que el mecanismo de biogénesis sea preciso. Si el proceso de biogénesis no fuera preciso, cambiaría la secuencia del miARN maduro y por lo tanto el reconocimiento del sitio blanco de los ARNm por este regulado impidiendo su regulación.

Existen diferencias durante la biogénesis y mecanismo de acción entre los miARNs en plantas y en animales. Pero además, además existen diferencias en las estructuras secundarias de los precursores de ambos reinos. En animales, existe una marcada homogeneidad en el tamaño de los precursores de miARNs, en cambio los precursores de miARNs de plantas son muy variables en forma y tamaño (Figura 1.4).

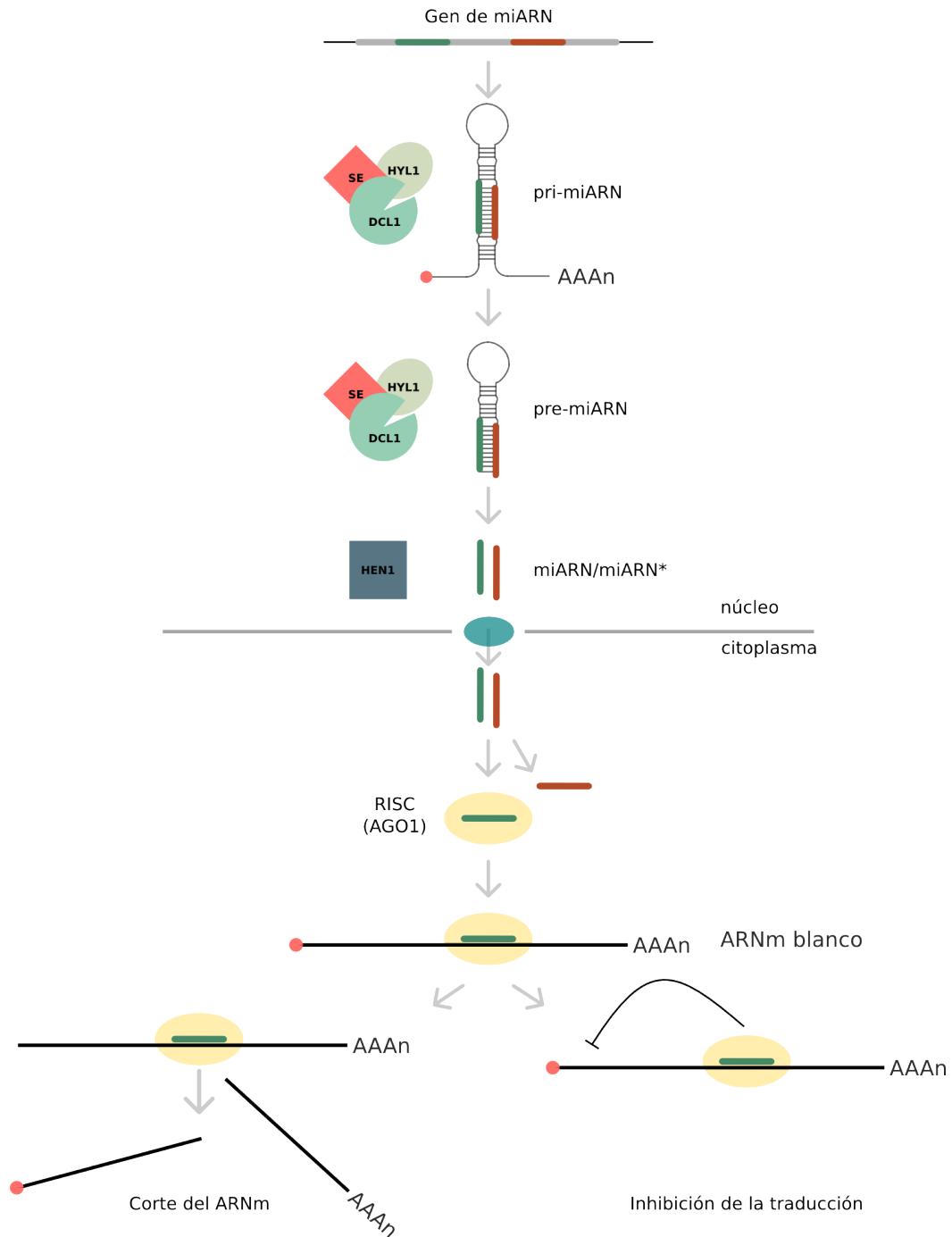


Figura 1.3: **Biogénesis y actividad de miARNs en plantas.** Los pri-miARNs son transcritos por la ARN polimerasa II y tienen extensa estructura secundaria en forma de tallo-burbuja (stem-loop). Estas estructuras son procesadas en el núcleo a través de la acción de una ribonucleasa llamada DICER LIKE 1 (DCL1) en asociación con el cofactor proteico de unión a ARN de doble hebra HYPONASTIC LEAVES 1 (HYL1) y la proteína SERRATE (SE). El producto generado a partir de los cortes llevados a cabo por DCL1, es un dúplex miARN-miARN\* que luego continúa siendo procesado por otros componentes enzimáticos hasta dar lugar al miARN maduro de 21 nt. Luego el miARN maduro se asocia con una proteína Argonauta (AGO), mayormente AGO1 y se incorpora dentro de un complejo de silenciamiento RISC. En el complejo RISC, el miARN principalmente induce el corte de los ARNm blanco, aunque algunos miARNs también regulan la expresión de sus genes blancos por represión traduccional.

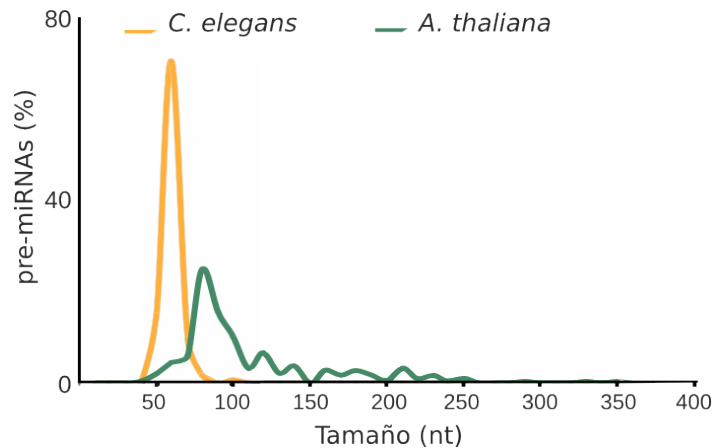


Figura 1.4: **Tamaños de precursores de miARNs.** En línea naranja se muestra la distribución de tamaños de pre-miARNs en *C. elegans* y en verde los pre-miARNs en *A. thaliana*. Figura adaptada de [21].

Trabajos previos que analizaron precursores de animales señalaron que un pri-miARN típico consiste en una estructura de tallo imperfecto secundado por segmentos inestables de ARNsh en sus extremos. El tallo imperfecto posee un tamaño de  $\sim 65$  nt en total que equivale a un dúplex de ARN de  $\sim 33$  pb [62]. A su vez señalaron que el pri-miARN puede ser dividido en 4 partes: un loop terminal, un dúplex miARN/miARN\*, un tallo inferior y secuencias flanqueantes simple hebra. Donde el dúplex miARN/miARN\* posee una longitud de  $\sim 22$  pb y el tallo inferior, que se encuentra debajo de la posición donde DROSHA produce el corte, posee una longitud aproximada de  $\sim 11$  pb [62]. Luego del primer corte llevado a cabo por DROSHA, el pre-miARN liberado está conformado por el dúplex miARN/miARN\* y el loop terminal.

Muchos precursores en plantas tienen un tallo de  $\sim 15$  nt debajo del dúplex miARN/miARN\* seguido por un loop interno, que sirve como una señal estructural de reconocimiento por la maquinaria de procesamiento [62, 94, 130, 151] (Figura 1.5). Sin embargo, este determinante de procesamiento no se encuentra en todos los precursores [94]. Además, la biogénesis de los miARNs conservados evolutivamente como ser el miR319 y miR159 comienzan con un corte al lado del loop terminal y continúa con 3 cortes adicionales en una dirección de burbuja a base hasta que finalmente el miARN es liberado [4, 23] (Figura 1.5). Se ha demostrado que otros precursores de plantas liberan otros ARNs pequeños además del miARN [75, 167], aunque no se conoce función biológica para estos otros miARNs. En principio, podrían ser una consecuencia secundaria del mecanismo de biogénesis del miR319.

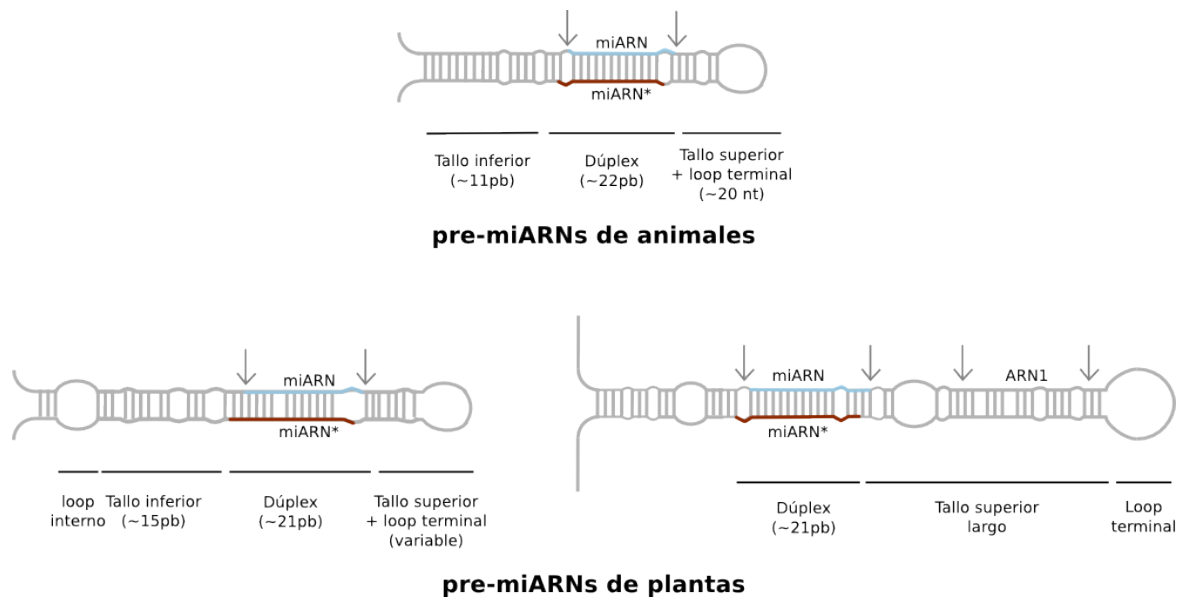


Figura 1.5: **Estructuras secundarias de precursores de miARNs.** Arriba se muestra la estructura secundaria de precursores típicos de animales. Abajo a la izquierda se muestra la estructura secundaria de precursores procesados de abajo en plantas. Abajo a la derecha se muestra la estructura secundaria de precursores que son procesados de arriba secuenciales en plantas. Las flechas indican los cortes necesarios para liberar el miARN maduro.

## 1.5. Regulación de la expresión génica por miARNs.

### 1.5.1. Regulación por corte del ARN blanco.

En animales existe un gran número de genes blanco mediado por miARNs y un ARNm puede estar regulado por varios miARNs, en cambio los miARNs en plantas regulan un número limitado de genes blanco [148].

El ARN pequeño guía al complejo RISC hacia una molécula de ARNm complementario. Luego del reconocimiento de ARNm blanco por complementariedad de bases, la proteína AGO1 del complejo RISC introduce un corte en un enlace fosfodiéster del ARNm. Este corte ocurre entre las posiciones 10 y 11 desde el extremo 5' del miARN, independientemente de la longitud del miARN [49, 84, 89, 107, 155].

Luego del corte mediado por el miARN, los fragmentos 3' son degradados mediante la actividad de la enzima citoplasmática 5'-3' EXORIBONUCLEASA4 (XRN4) en *En A. thaliana* [131]. Los fragmentos 5' también pueden ser degradados por el complejo denominado Exosoma, el cual está encargado de diferentes funciones de degradación y procesamiento

de ARNs [32]. En *Arabidopsis* la degradación del fragmento 5', puede ser acelerada por uridilación en el extremo 3' por la enzima una enzima denominada HESO1 [117].

### **1.5.2. Regulación de la traducción por miARNs.**

Los miARNs en animales, en general, son parcialmente complementarios a uno o más sitios presentes en la región 3' no traducida de los ARNm blancos [30, 46, 79] y esos ARNm raramente sufren el tipo de corte antes mencionado. Además, la limitada complementariedad de secuencia, permite que los miARNs de animales regulen generalmente la expresión de muchos genes blanco diferentes. Un mecanismo que involucra la inhibición de la traducción del ARNm blanco por el miARN explica la represión de la expresión de los blancos de miARNs en animales [46]. En otras ocasiones, los miARNs de animales disminuyen la vida media de los transcritos a los que se unen [59]. Existen algunos ejemplos de miARN de plantas también, donde se ha demostrado la existencia de un mecanismo de represión traduccional, además del corte del transcripto [11, 28, 34, 45, 77, 127].

### **1.5.3. Generación de ta-siRNAs y siRNAs secundarios.**

El corte de un transcripto por un miARN puede inducir la generación de ARN pequeños secundarios [5, 33, 37, 90, 100, 100]. Existe un mecanismo por el cual se genera un ARN pequeño específico en plantas que se denomina ta-siARN (trans-acting short-interfering RNA). Los genes que darán lugar a ta-siRNAs son transcritos que generalmente son denominados TAS. El transcripto es identificado y cortado por acción de un miARN que posee una secuencia complementaria al transcripto [5]. Los fragmentos de corte de un TAS se estabilizan y por la acción de ARN polimerasas dependientes de ARN, se convierten en un ARN de doble hebra. Luego el ARN doble hebra es procesado por una Ribonucleasa de tipo III, llamada DICER- LIKE4 (DCL4), generando ARNs pequeños de 21 nucleótidos, los ta-siARNs. Finalmente, estos ta-siARNs se incorporan a un complejo RISC y guían el corte de los transcritos de otros genes de manera similar a la acción de un miARN [5, 51, 154, 163].

## 1.6. Procesos biológicos regulados por miARNs en plantas.

Los procesos biológicos regulados por miARNs en plantas son muchos, y algunos miARNs desempeñan papeles importantes en el desarrollo, otros en la trasducción de señales hormonales, respuesta a estrés y respuesta a señales del ambiente [120, 121, 148].

De las 22 familias de miARNs conservados en plantas, 11 regulan factores de transcripción y la mayoría de ellos están involucrados en procesos de desarrollo o diferenciación celular [67] (Tabla 4.1). Por ejemplo, los genes blanco del miR172 pertenecen a la familia de factores de transcripción AP2 (APETALA2) que está involucrado en el tiempo de floración y en el desarrollo de la hoja [11, 34]. La familia del miR319 regula factores de transcripción TCP (TEOSINTE BRANCHED1, CYCLOIDEA, PCF1/2), la familia del miR165/166 factores de transcripción HD-ZIP III (HOMEODOMAIN-LEUCINE ZIPPER de clase III) y la familia del miR396 regulan factores de transcripción GRF (GROWTH-REGULATING FACTOR). Todas estas familias intervienen en el desarrollo de la hoja regulando distintos factores de transcripción [27, 107, 119]. El miR164, participa en el establecimiento del meristema apical del tallo, en el desarrollo de la raíz y definición de los bordes de los órganos mediante la regulación de miembros de la familia de factores de transcripción NAC (NAM, ATAF1/2 and CUC2) [78].

Por otra parte, entre los genes blanco que no corresponden a factores de transcripción, hay genes involucrados en diversos aspectos de la biología vegetal. Algunos codifican proteínas pertenecientes a la familia F-Box o enzimas E2 conjugantes de ubiquitina, las cuales están implicadas en la marcación de proteínas para la degradación por el proteosoma [120]. Otros genes con función conocida codifican para proteínas involucradas en la transducción de señales hormonales, o proteínas involucradas en el metabolismo energético, la respuesta a estrés o déficit de nutrientes.

Por otro lado, en especies en las que se han realizados estudios de ARN pequeños por técnicas de secuenciación de alto rendimiento, se han encontrado miARNs no conservados [38, 112, 119]. Los genes blanco regulados por estos miARNs tienen funciones más variables que los blanco de los miARNs conservados, no está clara la importancia biológica de estas interacciones [38].

En esta tesis nos enfocaremos principalmente en las familias de miARNs conservadas entre distintas especies de plantas (Tabla 4.1). Estas familias han sido estudiadas profundamente, por lo que su mayoría se encuentran ampliamente caracterizadas.



## 1.7. Predicción de genes blanco de miARNs.

### 1.7.1. Interacción de miARN-ARNm blanco.

Los miARNs reconocen secuencias complementarias en ARNm blancos provocando su corte o el arresto traduccional. Originalmente, se creía que los miARNs de plantas tenían complementariedad casi perfecta con los genes blanco regulados por ellos [17, 30, 85]. Sin embargo, en artículos más recientes se ha demostrado que existen genes blanco regulados por miARNs de distintas especies, con un mayor número de “mismatches” a lo que se creía previamente [53] (Figura 1.6).

Nos referimos como desajuste (utilizaremos el término en Inglés: “mismatches” el resto de la Tesis) a cualquier interacción entre pares de bases que no sean interacciones G-C o A-U. En el apareamiento de ARN a menudo se da una interacción no canónica denominada “wooble” que se forma cuando interacciona el par G-U. Este tipo de interacción, en muchos estudios es considerado un “mismatch” pero también puede ser considerado como un apareamiento. Otro tipo de interacción que se puede dar entre dos cadenas de ARN, son los denominados “bulges” que se forman cuando hay un nucleótido extra de un lado de la interacción que hace que una base en particular no interacciona con ninguna base enfrentada.

Teniendo en cuenta esto, vale mencionar que recientemente se han descrito genes blanco que tienen “bulges” de varios nucleótidos con sus miARNs que lo regulan [29]. Esto hace que la predicción de genes blanco regulados por miARNs en plantas no sea trivial y que se necesite información extra además de la interacción del par miARN-blanco.

Por otro lado en animales, la especificidad y la función de los miARNs están determinados por los nucleótidos 2 a 7 de la parte 5' del miARN maduro (llamada región “seed” o semilla del miARN) [83]. Dichos nucleótidos deben ser complementarios al ARNm blanco (Figura 1.6).

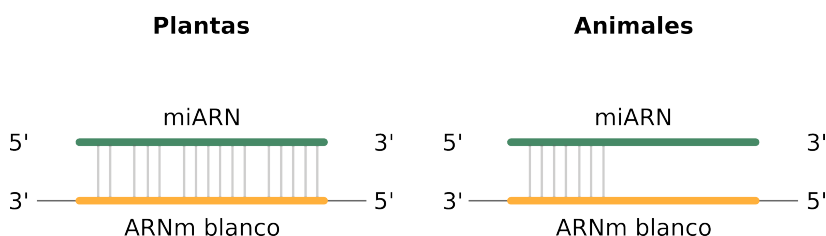


Figura 1.6: **Interacción de miARN-ARNm blanco.** A la izquierda se muestra una interacción del par miARN-ARNm blanco en plantas. A la derecha se muestra una interacción del par miARN-ARNm blanco en animales. Las líneas verticales grises representan bases apareadas, la ausencia de líneas representa “mismatches”.

### 1.7.2. Métodos para predicción de genes blanco de miARNs en animales.

Existen muchos métodos para la predicción de genes blanco de miARNs que fueron desarrollados por distintos grupos de investigación. Gran parte de ellos continúan manteniendo y actualizando sus algoritmos. El grupo de Cohen en el EMBL propuso el primer predictor de genes blanco de miARNs en 2003 [134] que luego fue actualizado en el 2005 [133]. TargetScan y TargetScanS fueron desarrollados por Bartel en el MIT y Burge en Cambridge [50, 57, 81, 82]. DIANA-microT, otra herramienta popular que fue creada por el grupo de Hatzigeorgiou, fue actualizada recientemente a la versión 5.0 [91, 92, 109]. El grupo de Rajewsky publicó su predictor de genes blanco en el 2005 y fue actualizado en el 2011 [9, 115].

Hay dos categorías de modelos predictivos: heurísticos y empíricos. Los modelos heurísticos utilizan algoritmos de detección que buscan posiciones a lo largo de la secuencia del ARNm y funciones de puntajes que filtran los blanco mediante la combinación de varios valores de entrada. Los primeros predictores aplican enfoques heurísticos, debido a la falta de una cantidad suficiente de datos para construir los modelos basados en el conocimiento empírico.

Los modelos predictivos utilizan entradas que se derivan del conocimiento de los detalles mecanísticos de las interacciones miARN-ARNm. La entrada más utilizada de los modelos predictivos, es la complementariedad de base entre el miARN y ARNm. Los miARNs de animales por lo general se unen al ARNm con sólo algunas posiciones que están apareadas [76], pero la complementariedad del apareamiento de bases en la región semilla (seed), que comprende los ocho primeros nucleótidos en el extremo 5' del miARN, es particularmente importante. Los parámetros de entradas de accesibilidad del sitio y de conservación evolutiva se utilizan para aumentar la especificidad. La mayoría de los predictores usan software existentes, como el Vienna RNA package [87], mFold [169], DINAMelt [93] y mFold [44], para el cálculo de la energía libre. El uso de la conservación evolutiva de los genes blanco de miARNs está motivada por la premisa de que las especies 'similares' deben compartir miARNs comunes y genes blanco comunes. Sin embargo, esto conduce a la omisión de los genes blanco no conservados [80, 139].

### 1.7.3. Métodos para predicción de genes blanco de miARNs en plantas.

Los algoritmos mencionados anteriormente son utilizados generalmente para predecir genes blanco de miARNs en animales. Algunos de ellos, como MiRanda, se utilizan tanto en animales como en plantas. Existen herramientas diseñadas específicamente para la predicción de genes blanco de miARNs en plantas. Algunas de ellas fueron implementadas como herramientas independientes, otras como servidores web y algunas como ambos.

La herramienta psRNATarget [40] utiliza un enfoque de programación dinámica, alineando las secuencias utilizando el algoritmo de Smith-Waterman modificado y aplicando el algoritmo 'RNAup'. Targetfinder [20] implementa un programa "FASTA" junto con un sistema de puntuación para "mismatches", "bulges" o "gaps" para alinear las secuencias. TAPIR [24] está integrado con dos opciones de búsqueda, el motor de búsqueda "FASTA" (para búsquedas rápidas), y el motor de búsqueda "RNA hybrid" (para obtener resultados precisos). La herramienta Target-align [153] también emplea el método de puntuación basado en el algoritmo de Smith-Waterman para predecir las complementariedades entre los miARNs y los RNAs mensajeros. Por último, la herramienta p-TAREF [65] implementa un modelo de regresión de máquinas de vectores de soporte (SVR en inglés "support vector regression") y utiliza la información de la "variación de la densidad de dinucleótido" alrededor del sitio blanco de un conjunto de datos de *A. thaliana*, *Oryza sativa*, *Medicago truncatula* y *Solanum lycopersicum*.

Además existen otras herramientas que son utilizadas para predecir genes blanco de miARNs pero que no fueron específicamente diseñadas para esto. Por ejemplo, la herramienta Web miRNA designer, WMD3 [106] es utilizada para el diseño de miARNs artificiales para silenciar la expresión de genes blanco específicos. Pero esta herramienta también es empleada para la predicción de genes blanco de miARNs en plantas. RNAHybrid [73] es una herramienta utilizada para buscar la energía mínima libre de un ARN largo y uno corto. La secuencia corta se hibrida con la parte más adecuada de la secuencia larga. La herramienta está destinada principalmente para la predicción de genes blanco de miARNs, pero no es su único uso.

Patmatch [158] que está integrado en "The Arabidopsis Information Resource" (TAIR)<sup>1</sup> permite realizar búsqueda de secuencias de nucleótidos o de péptidos corto, como dominios pequeños o motivos. Esta herramienta puede ser adaptada, para utilizarla como herramienta de búsqueda de genes blanco de miARNs en plantas.

---

<sup>1</sup><https://arabidopsis.org>

La mayoría de las herramientas específicas de plantas fueron desarrolladas para predecir genes blanco con una alta especificidad en el organismo modelo *A. thaliana*, si se utilizan los parámetros de predicciones óptimos. Y los sistemas de puntajes optimizados de Arabidopsis no se pueden utilizar como un umbral en el análisis de organismos no modelos [132]. Además estas herramientas en general detectan una gran cantidad de potenciales genes blanco (que incluyen los validados experimentalmente), pero su debilidad radica en que predicen una gran número de falsos positivos.

#### **1.7.4. Identificación de genes blanco regulados por miARNs mediante técnicas de secuenciación de alto rendimiento.**

Con los avances de los métodos de secuenciación de alto rendimiento o “next-generation sequencing”, la identificación sistemática de genes blanco de miARNs específicos en un tiempo corto ya es un hecho. Un método denominado PARE (Construction of Parallel Analysis of RNA Ends) [52]. La misma consiste en una combinación de la estrategia modificada 5' RACE, en donde se secuencian masivamente ARNs que carecen de 5' CAP. Esto implica que se detectan ARNm fragmentados, pero no los intactos. Además, el proceso de secuenciación permite identificar el sitio exacto de fragmentación o corte del ARNm. Por este motivo se suele referir a las bibliotecas PARE como el degradoma del ARN de las células.

Un elemento importante es que los miARNs y siARNs guían el corte de sus ARNm blanco entre las posiciones 10-11 del extremo 5' del ARN pequeño. Es decir, dada una interacción miARN/sitio blanco, uno puede predecir exactamente donde se producirá el corte del ARNm. Y si este corte realmente existe en las células, los fragmentos de ARNm detectados por PARE deberían estar enriquecidos en este lugar preciso del ARNm.

A partir de las bibliotecas de PARE, se han desarrollado diferentes herramientas como SeqTar [168], que es un método para la identificación de corte guiado por miARNs de degradoma de transcritos de plantas. Por otro lado, CleaveLand [3] es un pipeline computacional para la detección de cortes de miARNs en datos de degradoma. Esta herramienta toma como entrada secuencias de degradoma, ARNs pequeños y una base de datos de ARNm y devuelve los potenciales genes blanco de esos ARNs pequeños. sPARTA (small RNA-PARE target analyzer) [68] es un pipeline paralelizado, para el análisis integrado de miARNs en plantas y un conjunto de ARNm. Este pipeline incluye un software de identificación de nuevos genes blanco regulados por miARNs y utiliza datos de PARE.

Por otro lado, la herramienta PAREsnip permite realizar búsqueda de potenciales genes blanco de todos los ARN pequeños obtenidos de técnicas de secuenciación de alto rendimiento. Mediante la búsqueda de genes blanco de todo el “sRNAome” se puede facilitar identificación de genes blanco de ARN pequeños a grande escala.

# Objetivos.

## 2.1. Objetivo general.

Las interacciones ARN-ARN son esenciales durante la regulación de la expresión génica por miARNs, y en los procesos de silenciamiento génico en general. La identificación de los genes blanco por miARNs se basa en la interacción de dos moléculas de ARN donde la interacción no es necesariamente perfecta. En principio, se puede hipotetizar que cierto número de bases desapareadas en determinadas posiciones permitirán la actividad del ARN pequeño en el contexto de la interacción miARN-gen blanco.

Por otro lado, la secuencia exacta del miARN depende del procesamiento de precursores más largos con extensa estructura secundaria. Este proceso es preciso en plantas, pese a que los precursores de miARNs son variables en secuencia, forma y tamaño.

El primer objetivo de esta Tesis se basa en diseñar estrategias para identificar genes regulados en miARN en plantas, tratando de seleccionar aquellas interacciones que tengan una mayor probabilidad de tener un impacto significativo en la biología de las plantas. En este proceso hemos elegido una combinación de herramientas bioinformáticas, incluyendo la conservación de la interacción miARN-gen blanco en distintas especies.

Un segundo objetivo busca encontrar patrones comunes y dominios de ARN que sirvan de guía durante el proceso de biogénesis de miARNs. Aquí también hemos elegido una combinación de elementos incluyendo el análisis de experimentos de secuenciación de alto rendimiento y el análisis de la conservación evolutiva de precursores de miARNs.

Se espera que los resultados de esta Tesis sirvan no solo para alcanzar los objetivos de investigación planteados sino también para promover el desarrollo de la Bioinformática como una disciplina que brinda una oportunidad única para que, a partir de investigaciones en las ciencias básicas, pueda hallarse el camino hacia el desarrollo de aplicaciones de interés estratégico para el país.

## **2.2. Objetivos específicos.**

1. Identificar genes regulados por miARNs en plantas.
  - Diseñar una estrategia para la identificación de genes blanco regulados por miARNs en plantas, basado en la conservación evolutiva del par miARN-gen blanco.
  - Desarrollar una herramienta web para la predicción de genes blanco de miARNs en diferentes especies de plantas.
2. Estudiar la biogénesis de los miARNs en plantas.
  - Desarrollar herramientas que permitan el análisis de los intermediarios de procesamiento de miARNs en plantas a partir de bibliotecas de secuenciación masiva de ARN.
  - Identificar y caracterizar de precursores de miARNs en distintas especies que tengan mecanismos de procesamiento distintos.
  - Caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento determinados previamente.

# Materiales y Métodos.

## 3.1. Métodos utilizados para la predicción de genes regulados por miARNs en plantas.

En la primer parte de esta tesis diseñamos una estrategia para la identificación de genes blanco regulados por miARNs basado en la conservación evolutiva del par miARN-gen blanco. La metodología aplicada es la siguiente.

### 3.1.1. MiARN consensos.

Las 22 familias de miARNs conservadas en angiospermas fueron consideradas para esta parte del trabajo [14, 48]. MiR319 y miR159 que codifican para miARNs similares, fueron considerados como familias diferentes ya que regulan a genes blanco distintos [108]. Consideramos todos los miembros de estas familia, obtenidos de miRBase<sup>2</sup>, pertenecientes a *A. thaliana*, *Populus trichocarpa* y *Oryza Sativa*. Variaciones en las posiciones 1, 20 y 21 son muy comunes en las familias de miARNs [42]. Por esto, definimos como secuencia consenso, a las secuencias más comunes (posiciones 2-19) de distintos miembros de cada familia (tabla 4.1).

### 3.1.2. Predicción de genes regulados por miARNs.

#### Conjunto de datos de plantas.

Los datos de las secuencias pertenecen a bibliotecas extraídas de “Gene Index Project”<sup>3</sup>, que consiste en una base de datos de ESTs ensamblados. Seleccionamos un conjunto de datos pertenecientes a Angiospermas. Además, utilizamos secuencias de ARNm completos de *A.*

---

<sup>2</sup><http://mirbase.org>

<sup>3</sup><http://compbio.dfci.harvard.edu/tgi/>



*thaliana*<sup>4</sup> y *Oryza Sativa*<sup>5</sup> (ver tabla 4.2). La búsqueda la realizamos utilizando PatMatch [158], que es un programa de búsqueda de patrones de nucleótidos cortos o péptidos. El programa puede ser usado para encontrar coincidencias con un patrón de secuencia específico y permite el uso de códigos de secuencias ambiguas y expresiones regulares y por esto se puede utilizar la búsqueda con “mismatches”, inserciones y deleciones. Realizamos la primera búsqueda de potenciales genes blanco permitiendo tres “mismatches” con las secuencias consensos, mientras que las interacciones G:U y los bulges fueron considerados “mismatches”. Para realizar el alineamiento del par miARN-gen blanco, desarrollamos una versión modificada del algoritmo de programación dinámica Needleman-Wunsch [101], utilizando el lenguaje Perl<sup>6</sup>. Además, desarrollamos scripts para integrar los módulos de Blastx [8] utilizando el proteoma de Arabidopsis y el módulo RNAhybrid [54] que es una herramienta que permite encontrar la menor energía libre de hibridación (MFE) de dos secuencias de ARN.

### Filtros.

Las secuencias candidatas fueron etiquetadas con el identificador del locus (locus ID) con mejor puntuación (best hit) en *A. thaliana*, utilizando el módulo de Blastx (Corte del evalue de  $10e^{-5}$ ). De este modo, genes blanco de distintas especies que tenían la misma etiqueta fueron agrupados juntos, ya que tendrían el mismo homólogo en *A. thaliana*. El filtro de conservación evolutiva hace referencia al número mínimo de especies donde la misma etiqueta estaba presente para un miARN particular. El filtro empírico está basado en trabajos previos [127] y hace referencia a la energía de interacción MFE (mínima energía libre de hibridación de al menos 72 % del apareamiento perfecto). El otro filtro empírico requiere que entre el par miARN-gen blanco, solamente está permitido un “mismatch” entre la posición 2 y la 12 del miARN (1-11 de nuestra búsqueda modificada con las secuencias consenso).

### Controles.

Como control, realizamos las búsquedas del mismo modo que lo hicimos para los miARNs conservados, pero utilizando secuencia al azar. Para cada miARN conservado, generamos 20 secuencias al azar (scramble) dividiendo las secuencias originales de a dinucleótidos y luego generando nuevas secuencias al azar conservando esa composición

---

<sup>4</sup><http://arabidopsis.org>

<sup>5</sup><http://rice.plantbiology.msu.edu>

<sup>6</sup><http://perl.org>

de los di-nucleótidos como fue descrito previamente [66]. De estas 20 secuencias al azar, elegimos las 10 que tenían el número más similar del total de genes blanco para el miARN real correspondiente. La relación señal/ruido fue calculada como el cociente entre el número de genes blanco para los miARNs y el número de genes blanco del promedio obtenido para las secuencias al azar. Como un control adicional, seleccionamos dos miARNs que no están conservados durante la evolución, que son el miR158 y el miR173.

### **3.2. ComTAR: una herramienta para la predicción de genes blanco regulados por miARNs en plantas.**

A partir de los resultados positivos obtenidos de la estrategia descrita anteriormente, decidimos desarrollar una herramienta web y dejarla disponible para la comunidad científica denominada comTAR que está disponible en un sub-dominio de la página web institucional del IBR: <http://rnabiology.ibr-conicet.gov.ar/comtar>.

#### **3.2.1. MiARN y transcriptos.**

Como las secuencias del maduro del miARN puede variar en distintas especies, especialmente en la posición 1, 20 y 21 ([35], utilizamos secuencias del 2-19 (18nt) para realizar las búsquedas. Como además existen variaciones en las secuencias en los distintos miARNs de las mismas familias, utilizamos la más representativa teniendo en cuenta los genomas de *Arabidopsis*, álamo y arroz. De este modo comTAR contiene datos pre-calculados, de potenciales genes blanco para 22 miARNs conservados en plantas (ver tabla 4.1) donde el usuario puede navegar los resultados y cambiar los parámetros de entrada. Además, el usuario puede realizar la búsqueda de nuevos ARNs pequeños teniendo en cuenta esta consideración. El cálculo se hace en el cluster del CCT-Rosario y los datos se obtienen luego de unas horas. Como la herramienta web la realizamos tiempo después de haber hecho la estrategia para predicción de genes blanco, utilizamos una nueva base de datos más actualizada y completa denominada Phytozome<sup>7</sup> [55]. La misma corresponde a secuencias de transcriptos de plantas formado por archivos de nucleótidos en formato FASTA de transcriptos de ARNm (UTR, exones) con variantes de splicing.

---

<sup>7</sup><http://phytozome.jgi.doe.gov>

### 3.2.2. Búsqueda de genes blanco.

La búsqueda de genes blanco la realizamos de la misma manera que la descrita anteriormente con algunos cambios. Además de actualizar la base de datos y utilizar la de Phytozome, actualizamos la base de datos de *A. thaliana* por la del TAIR10. Las secuencias candidatas fueron etiquetadas con el mejor hit del locus ID del Arabidopsis TAIR10, utilizando los archivos de anotación de Phytozome, y lo utilizamos como “TAG” (etiqueta). Por último, cada TAG de Arabidopsis fue indexado con una breve descripción funcional y computacional obtenida del TAIR10 y los genes blanco candidatos fueron agrupados por familias teniendo en cuenta la clasificación de familias del TAIR10.

### 3.2.3. Herramienta web y almacenamiento de datos.

ComTAR fue diseñado como una aplicación web con un framework open-source en PHP denominado Codeigniter para la interfaz gráfica, pero el análisis está basado en un back-end escrito en Perl. Los datos que surgen de ese análisis fueron almacenados en una base de datos en MySQL<sup>8</sup>. El back-end es el encargado de realizar la búsqueda de secuencias y además ahí es donde se integraron las herramientas y scripts para aumentar la especificidad y sensibilidad de comTAR. También el back-end es el encargado de generar los resultados finales. Mientras el front-end es el responsable de mostrar los resultados (Figura 3.1). El TAG del mejor hit en Arabidopsis es el que determina el número de especies donde un gen blanco está presente, y el número mínimo de especie es un parámetro que es definido por el usuario.

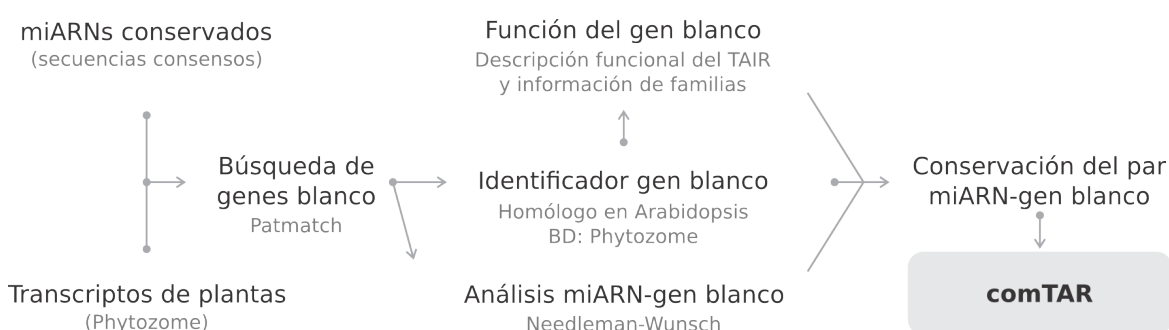


Figura 3.1: comTAR. Diagrama de flujo que describe la herramienta

<sup>8</sup><http://mysql.com>

### 3.3. Análisis bioinformático de los datos obtenidos de la secuenciación por la técnica de SPARE.

Una vez obtenidos los datos crudos de las secuencias se procedió a hacer el procesamiento de los datos siguiendo los siguientes pasos:

- Comprobar la calidad de las secuencias.
- Remover secuencias adaptadoras (trimming).
- Agrupar secuencias iguales y determinar abundancia (conversión de archivos fastq file a archivos tag-count).
- Mapear las secuencias únicas con las secuencias de precursores (sub-genoma definido).
- Combinar los resultados del mapeo con la abundancia de cada secuencia.
- Analizar y comparar los intermediarios de procesamiento para plantas silvestres y mutantes de procesamiento.

El control de calidad de las secuencias fue realizado con la herramienta FastQC<sup>9</sup> (Versión v0.10.1). Luego removimos las secuencias adaptadores con la herramienta cutadapt<sup>10</sup> (Versión 1.9). A continuación, mediante scripts desarrollados por nosotros utilizando comandos de bash en Linux, agrupamos las secuencias y determinamos la abundancia de cada una.

### 3.4. Desarrollo de una herramienta intuitiva para el análisis y visualización de los datos obtenidos por la técnica de SPARE.

La herramienta fue desarrollada para facilitar el análisis de la identificación de los intermediarios de procesamiento en bibliotecas de SPARE. En total se obtuvieron cerca de 150 millones de secuencias sumando todas las bibliotecas. Esos fragmentos son agrupados y mapeados contra las secuencias únicas de los precursores de *A. thaliana*. Nos quedamos con 7158 secuencias únicas, que luego fueron filtradas por largo de la secuencia, quedándonos con las de largo mayor o igual a 30 nt (Figura 3.2).

<sup>9</sup><http://bioinformatics.babraham.ac.uk/projects/fastqc/>

<sup>10</sup><http://cutadapt.readthedocs.org/>

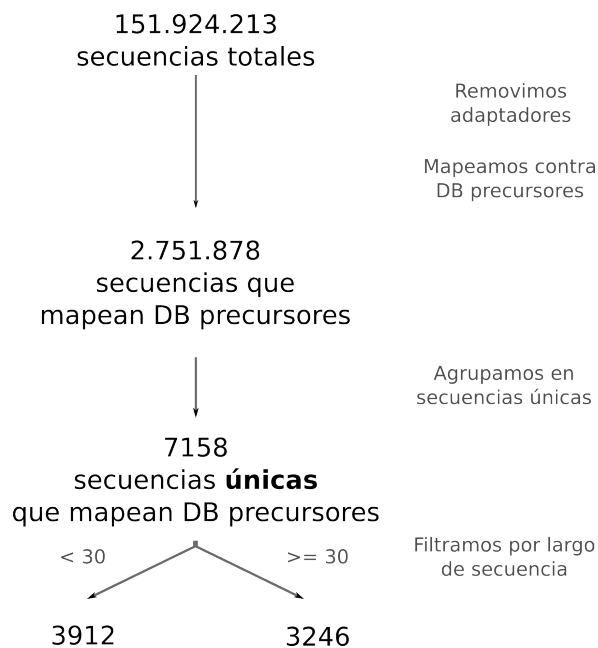


Figura 3.2: **Análisis informático utilizado para procesar los datos de SPARE.** Las secuencias únicas mapeadas contra los precursores son las que se utilizan en la herramienta para analizar los intermediarios de procesamiento.

Esa información ya procesada fue almacenada en una base de datos, y es la que utilizamos en la herramienta para analizar y comparar los intermediarios de procesamiento para plantas silvestres y para las mutantes de procesamiento.

### 3.5. Análisis estructural de precursores de miARNs en plantas.

Obtuvimos las estructuras secundarias para cada precursor calculada a partir de la herramienta Mfold [169] con los parámetros por default a 37 °C de temperatura. El lado proximal del dúplex miARN/miARN\* fue definido como la posición +1. Analizamos la estructura secundaria y consideremos las posiciones que había un match como un 0, mientras que bulges y “mismatches” los consideramos como 1. Además, hicimos un promedio para todos los precursores siguiendo la misma estrategia. Implementamos un pipeline bioinformático utilizando “in-house” scripts y datos públicos de miRBase, para asistir con el análisis de las bibliotecas de secuenciación masiva. Las secuencias de los ARN pequeños fueron obtenidas de la base de datos de nueva generación de Arabidopsis<sup>11</sup> [68] y de la base de datos de miRBase [72].

<sup>11</sup><https://http://mpss.udel.edu/>

Los datos para el primer análisis de SPARE fueron realizados por Nicolás Bologna y Arnaldo Schapire sobre plantas silvestres y mutantes *fiery*. Los mismos están accesibles mediante el NCBI Gene Expression Omnibus (GEO<sup>12</sup> con el código de acceso GSE46429.

Además, analizamos datos de nuevas bibliotecas de SPARE realizadas por Belén Moro, para precursores de *A. thaliana* de plantas silvestres (Col-0) y en *hyll*, *se* y *fiery*.

### **3.6. Estrategia bioinformática para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento.**

#### **Búsqueda de ortólogos y extensión de la secuencia.**

De la base de datos biológica miRBase (release 19), obtuvimos las secuencias de 92 precursores que incluye cada miembro de las familias de *A. thaliana* conservados en dicotiledóneas. Además, obtuvimos las secuencias de 117 precursores que incluye cada miembro de las familias de *O. Sativa* conservados en monocotiledóneas. Con estas secuencias de *A. thaliana* y de *O. Sativa*, buscamos los ortólogos en dicotiledóneas y monocotiledóneas respectivamente. Esto lo hicimos utilizando la técnica de Blast recíproco, mediante un script desarrollado por nosotros que utiliza el blast del paquete de BLAST+ del NCBI (versión 2.2.25). El Blast recíproco es un método computacional para predecir ortólogos putativos. Se realiza tomando un gen de interés y luego se realiza un Blast de dicho gen sobre una base de datos de un organismo donde se quiere buscar el ortólogo. Luego, se toma el gen que se obtiene con mejor puntuación por Blast, y se utiliza para hacer un Blast sobre una base de datos del mismo organismo del gen de interés inicial. Si este Blast devuelve con el mejor puntaje al mismo gen inicial, entonces estos dos genes son considerados ortólogos putativos.

Los datos genómicos para la búsqueda de ortólogos, fueron extraídos de la base de datos de Phytozome y utilizamos 30 especies de dicotiledóneas y 6 especies de monocotiledóneas (Figura 3.3). Además, extendimos las secuencias obteniendo más bases de las que originalmente están definidas en miRBase y empezamos nuestro análisis con una definición arbitraria de los precursores de plantas incluyendo 150 nt fuera del par miARN/miARN\*. Esta definición se basa en que todos los determinantes estructurales validados experimentalmente necesarios para el procesamiento de miARNs están dentro de este rango.

---

<sup>12</sup><http://ncbi.nlm.nih.gov/geo>

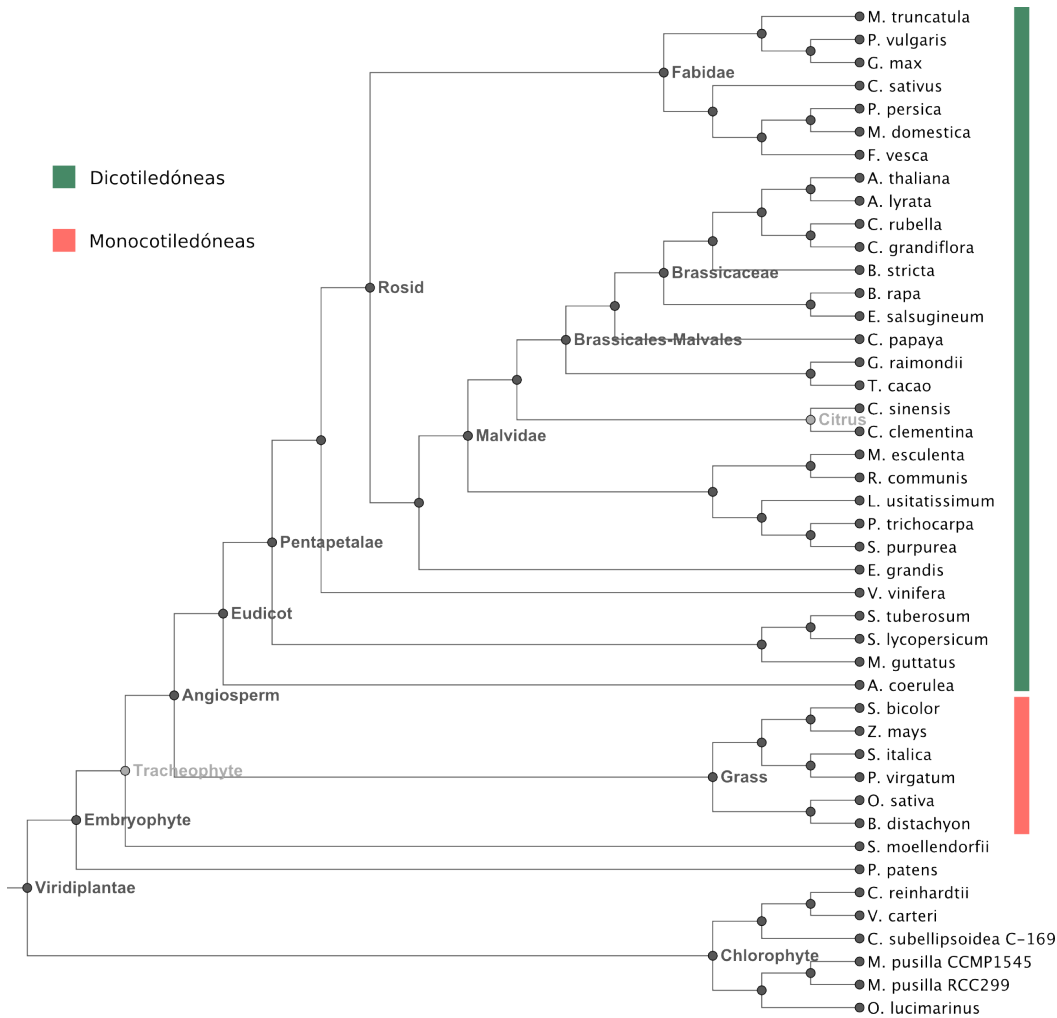


Figura 3.3: Especies almacenadas en Phytozome.

### Alineamientos de precursores y determinación de estructura secundaria.

Los alineamientos de precursores en base a su secuencia primaria se realizaron utilizando la herramienta T-coffee (versión 11.00.8cbe486) [103] como se muestra a continuación :

---

```
t_coffee -in=miR.fasta -mode=regular -method=slow_pair -output=fasta_aln -quiet=stdout -out=miR.aln
```

---

Por defecto, T-coffee colorea los alineamientos en base a un puntaje de las bibliotecas construida a partir de distintos algoritmos de alineamiento. Como a nosotros nos interesa poder visualizar las diferencias en base a la conservación de los precursores, utilizamos T-Coffee para colorear el alineamiento en base a el nivel de conservación

---

```
t_coffee -other_pg seq_reformat -in miR.aln -in3 miR.aln -action +3evaluate idmat -out=miR_T_Coffee.pdf
```

---

Los alineamientos de precursores, considerando su estructura secundaria, fueron realizados con la herramienta R-coffee [146] como se muestra a continuación.

---

```
t_coffee -in=miR.aln -mode=rcoffee -method=slow_pair -output=score_pdf score_ascii -run_name=miR_R_Coffee
```

---

Y para colorear los alineamientos en base a su conservación, utilizamos el siguiente comando:

---

```
t_coffee -other_pg seq_reformat -in miR.aln -in3 miR.aln -action +3evaluate idmat -out= miR.aln_R_Coffee.pdf
```

---

El plegamiento de los precursores para determinar la estructura secundaria lo realizamos utilizando el paquete RNAfold (versión 2.1.1) [87].

### Búsqueda de motivos conservados.

La búsqueda de motivos conservados la realizamos utilizando la herramienta MEME (versión 4.10.1) [16]. Hicimos dos búsquedas de motivos, la primera para encontrar como máximo 2 motivos conservados en los precursores, de alrededor de 20nt. Utilizamos el comando y los parámetros como se muestra a continuación:

---

```
meme miR.fasta -dna -oc miR -nostatus -time 18000 -mod zoops -nmotifs 2 -minw 20 -maxw 23 -revcomp
```

---

Y luego hicimos otra búsqueda de motivos, pero esta vez con tamaño más variable (de 5 a 50 nt) y permitimos encontrar hasta 10 motivos con un umbral de corte por e-value. Esto lo realizamos para encontrar otros motivos conservados además de los ARN pequeños. Utilizamos el comando y los parámetros como se muestra a continuación:

---

```
meme miR.fasta -dna -oc miR -nostatus -time 18000 -mod anr -nmotifs 10 -minw 5 -maxw 50 -revcomp -evt 1e-3
```

---



### **Representación gráfica de los precursores.**

En base a la información de los alineamientos y estructuras secundarias, representamos los datos utilizando el paquete Circos [74]. Por simplicidad en la representación de los precursores, eliminamos los gaps y sólo están representadas las bases dentro del precursor. El archivo de configuración principal se muestra en el anexo en 9.1 y los archivos de configuración del histograma y las relaciones se muestran en el anexo en 9.2 y 9.3 respectivamente. Los archivos karyotype.txt, highlight.txt y link.txt son distintos para cada precursor en particular y se realizaron mediante un script, con el formato que se especifica en <http://circos.ca/>.

### **3.7. Código fuente de los scripts y herramientas desarrolladas.**

Para todo este trabajo de Tesis se utilizaron la mayoría de programas de Software libre y varios de ellos de código abierto en un entorno Linux. Algunos de los programas, scripts y algoritmos desarrollados en esta Tesis están disponibles para su uso en los siguientes repositorios:

- La herramienta para la predicción de genes blanco regulados por miARNs en plantas comTAR, se encuentra alojada en [https://github.com/ucielp/cluster\\_comtar](https://github.com/ucielp/cluster_comtar).
- La herramienta para el análisis y visualización de los datos obtenidos por la técnica de SPARE, se encuentra alojada en [https://github.com/ucielp/SPARE\\_tool](https://github.com/ucielp/SPARE_tool).
- Los scripts para caracterizar la relación entre la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento, se encuentran alojados en [https://github.com/ucielp/miRNAs\\_evolution](https://github.com/ucielp/miRNAs_evolution).

### **3.8. Anexo: Técnicas de Análisis Molecular y bioquímico relacionadas al trabajo de Tesis.**

Si bien esta Tesis es de carácter bioinformático, el trabajo realizado también fue interdisciplinario interactuando con otros miembros del grupo para concretar objetivos más generales. Parte de este trabajo en biología molecular y bioquímica fue realizado para validar las predicciones aquí descriptas. Alternativamente, el trabajo aquí presentado trató del análisis informático de datos genómicos como los de las bibliotecas SPARE realizadas experimentalmente por otros miembros del grupo. Si bien estas secciones no son parte de los

materiales y métodos propiamente realizados, consideramos relevantes mencionarlos ya que contribuyen a un entendimiento general del problema que aquí se trata.

### **3.9. Obtención de Arabidopsis transformadas.**

#### **3.9.1. Ecotipos utilizados y condiciones de crecimiento.**

Las plantas de *A. thaliana* utilizadas para los experimentos en esta parte del trabajo corresponden a el ecotipo Columbia-0 Col-0. Las plantas fueron cultivadas en una cámara de crecimiento con un régimen de 16 h de luz ( $100 \mu\text{E.m.}^{-2}\text{s}^{-1}$ ) y 8 h de oscuridad (condición día largo). La temperatura de crecimiento fue de 23°C durante el ciclo luz/oscuridad, mientras que la humedad fue mantenida en 65 % de humedad relativa. Las plantas fueron regadas 2 veces por semana con agua. Para el crecimiento directo en tierra, las semillas fueron estratificadas a 4°C por 2 días en tubos de microcentrífuga con 1ml de 0,1 % (p/v) agar, y luego sembradas en tierra. Las plantas de *Nicotiana tabacum* (cv Petit Havana) fueron crecidas en condición día largo durante 8 semanas y la segunda hoja fue utilizada para el análisis de ARN.

#### **3.9.2. Transformación de Arabidopsis.**

La transformación de plantas de Arabidopsis se realizó mediante la técnica de “floral-dip”. Se cultivó las cepas de *A. tumefaciens* ASE transformadas con los distintos plásmidos binarios en 100 ml de LB conteniendo 50  $\mu\text{g/ml}$  de espectinomicina, 50  $\mu\text{g/ml}$  de kanamicina y 25  $\mu\text{g/mg}$  de cloranfenicol a 28 °C hasta llegar a una DO600 cercana a 2. Se cosechó las células por centrifugación a 5.000 g durante 10 min a temperatura ambiente y se resuspendió en un medio con 5 % (p/v) sacarosa y 200  $\mu\text{l/l}$  de Silwet L-77. Las inflorescencias de plantas de Arabidopsis regadas el día anterior se sumergieron en la suspensión de células durante 30 segundos. Luego de escurrir el exceso de medio de cultivo, se colocaron las plantas en posición horizontal dentro de bolsas plásticas para conservar la humedad. Luego de 24 hs se sacaron de las bolsas y se prosiguió con el cultivo de las plantas en la cámara de crecimiento hasta la maduración de las semillas.

### 3.9.3. Recolección y esterilización de semillas.

La recolección de semillas se realizó a partir de plantas que finalizaron su ciclo de vida y que presentaban semillas maduras. Se retiró las plantas del pote de tierra y frotando las mismas con las manos se dejó caer las semillas sobre una hoja de papel. Se eliminaron los restos de silicua y hojas por tamizado. Las semillas se almacenaron en bolsas de papel en un recipiente seco hasta su posterior utilización.

Para la esterilización de las semillas, las mismas se incubaron con una solución de 70 % (v/v) etanol, 0,1 % (v/v) tritón-X-100, durante 8 min. Luego se realizó un enjuague con 96 % (v/v) etanol y finalmente se las dejó secar en el flujo laminar.

La manipulación del material vegetal en condiciones de esterilidad se realizó en una cabina de flujo laminar.

### 3.9.4. Selección de transformantes.

Para la selección de plantas transformadas, las semillas esterilizadas se sembraron en placas de Petri con medio MS-agar suplementado con 25  $\mu\text{g/ml}$  de kanamicina. Las placas se mantuvieron en oscuridad a 4 °C por 48 hs. Transcurrido este período se transfirieron las placas a la cámara de cultivo y en estas condiciones las plantas se dejaron crecer 5-10 días. Luego se analizaron las placas y aquellas plantas que presentaban un fenotipo de resistencia al agente de selección, evidenciado por el desarrollo de la raíz primaria y del primer par de hojas verdaderas, se transfirieron a potes con tierra. Los potes se ubicaron en bandejas con agua las cuales se cubrieron con papel de nylon transparente, de modo de evitar la pérdida de humedad. En estas condiciones se cultivaron en la cámara durante 2 días, momento en el que se retiró el nylon y se continuó con su cultivo.

### 3.9.5. Transformación nuclear de tabaco.

Las plantas transgénicas de tabaco fueron preparadas por el método de transformación de discos de hojas por co-cultivo con *A. tumefaciens*. Se cultivó la cepa de *A. tumefaciens* GV3101 pMP90 transformada con los distintos plásmidos binarios en 10 ml de 2YT conteniendo los antibióticos necesarios (100  $\mu\text{g ml}^{-1}$  rifampicina, 50  $\mu\text{g ml}^{-1}$  gentamicina, 60  $\mu\text{g ml}^{-1}$  cloranfenicol) a 28 °C hasta llegar a una DO600 entre 0,5 y 1. Se utilizaron plantas de tabaco (*N. tabacum cv Petit Havana*) crecidas en medio de cultivo sólido de Murashige-Skoog con 3 % (p/v) sacarosa (MS-0) en condiciones estériles. Las hojas fueron cortadas en piezas

de aproximadamente 1  $cm^2$  usando un bisturí estéril. Las porciones de hoja fueron colocadas en una caja de Petri conteniendo 100 ml de medio MS líquido, al que se agregaron 10 ml del cultivo de *A. tumefaciens*, y se incubó entre 15 y 30 min. El exceso de líquido de las porciones de hoja fue eliminado con papel de filtro estéril y luego se las colocó en placas de Petri conteniendo medio MS-10 (MS-0 suplementado con 1 mg l<sup>-1</sup> kinetina y 0,1 mg l<sup>-1</sup> ácido naftalenacético). Tras 4 días de incubación a 28 °C, los discos de hoja fueron cambiados a cajas de Petri con medio MS-10 suplementado con 100  $\mu$ g l<sup>-1</sup> kanamicina, para seleccionar células vegetales transformadas, y 100  $\mu$ g ml<sup>-1</sup> cefotaxima, para eliminar el *Agrobacterium*. Luego de 3 a 4 semanas de crecimiento en cámara de cultivo, se generaron vástagos que fueron transplantados a medio MS-0 suplementado con los mismos antibióticos, para que desarrollen raíces y continúen con su crecimiento. Después de 6 a 8 semanas de crecimiento, las plántulas que desarrollaron raíces fueron transplantadas a tierra y aproximadamente 12 semanas después se cosecharon las semillas.

### **3.10. Cuantificación del nivel de expresión génica.**

#### **3.10.1. Extracción de ARN.**

La extracción de ARN de tejido vegetal de *Arabidopsis* se realizó utilizando el reactivo TRIzol (Invitrogen). La recolección de las muestras de tejido se realizó en tubos de microcentrífuga que inmediatamente fueron sumergidos en N<sub>2</sub> líquido. El material fue reducido a un polvo fino utilizando un pilón. A continuación, sobre este polvo se adicionó TRIzol (1 ml de reactivo para un máximo de 100 mg de tejido) y se agitó de modo de resuspender el tejido mortereado. El homogenado se centrifugó a 12000 g a 4 °C durante 10 min y finalmente se transfirió el sobrenadante a un tubo de microcentrífuga nuevo.

A continuación se agregó 0,2 volúmenes de cloroformo por cada volumen de TRIzol original. Esta mezcla se invirtió vigorosamente por 15 segundos y se centrifugó a 12000 g a 4 °C durante 15 min. Se transfirió la fase superior a un nuevo tubo de microcentrífuga y se precipitó el ARN mediante el agregado de un volumen de isopropanol. Esta mezcla se incubó por 2 hs a -20 °C. Luego se centrifugó a 12000 g a 4 °C durante 10 min. Finalizada la centrifugación se descartó el sobrenadante.

Se lavó el precipitado de ARN mediante el agregado de 1 ml de 70 % (v/v) etanol frío y agitación con vortex. Luego se centrifugó a 7500 g a 4 °C durante 5 min, descartando el

sobrenadante una vez finalizada la centrifugación. Este paso de lavado se repitió una vez más.

El precipitado de ARN obtenido se secó en estufa 37 °C por 10 min y luego se resuspendió en 50  $\mu$ l de agua Milli-Q esteril.

### **3.10.2. Cuantificación y chequeo de la integridad del ARN purificado.**

Se determinó la absorbancia a 230, 260 y 280 nm. Se estimó la pureza de la preparación a partir de la relación de las medidas de absorbancia a Abs260/ Abs230 y Abs260/ Abs280. La integridad del ARN purificado se determinó mediante la electroforesis en geles de 1,5 % (p/v) agarosa de 5  $\mu$ l del ARN preparado. Las bandas de ARN ribosomal se visualizaron por tinción de los geles con bromuro de etidio.

### **3.10.3. Tratamiento del ARN preparado con ADNasa.**

Se preparó la siguiente mezcla de reacción: En un volumen final de reacción de 20  $\mu$ l se adicionó: 0,5 a 1  $\mu$ g de ARN total, 2  $\mu$ l de buffer “RQ1 RNase-Free DNase” (Promega), 1 U “RQ1 RNase-Free DNase” (Promega) y agua Milli-Q para completar el volumen de reacción.

La mezcla de reacción se incubó 30 min a 37 °C. Luego se inactivó la ADNasa mediante la adición al tubo de reacción de 1  $\mu$ l de “DNase Stop Solution” (Promega) e incubación por 10 min a 65 °C. De los 21  $\mu$ l de reacción finales, 12  $\mu$ l se utilizaron para sintetizar ADN complementario (ADNc). Los 9  $\mu$ l restantes se utilizaron como control negativo en la PCR en tiempo real.

### **3.10.4. Retrotranscripción (RT).**

La síntesis del ADNc a partir del ARN preparado se llevó cabo según el siguiente protocolo. En una primera etapa se preparó la siguiente mezcla: en un volumen final de 13,5  $\mu$ l se adicionó 0,25  $\mu$ g de oligo dT, 12  $\mu$ l de ARN tratado con ADNasa, dNTPs a una concentración final de 0,4 mM cada uno. El volumen de la mezcla se completó con agua Milli-Q.

La mezcla de reacción se incubó durante 5 min a 65 °C y a continuación en hielo durante al menos un minuto. Luego se realizó una centrifugación rápida y se adicionó a cada tubo de reacción 4  $\mu$ l de “5X First Strand Buffer”, 1  $\mu$ l de 0,1 M DTT, 1  $\mu$ l “Rnase

OUT Recombinant Ribonuclease Inhibitor” (Invitrogen) y 100 U de “SuperScript III Reverse Transcriptase” (Invitrogen).

Se mezcló por inversión y se incubó 60 min a 50 °C. Luego se realizó una centrifugación rápida y se inactivó la reacción mediante la incubación a 70 °C durante 15 min. Finalmente el ADNc a usar como molde en la reacciones de amplificación se diluyó al menos 40 veces en agua Milli-Q.

### 3.10.5. Reacción en cadena de la polimerasa en tiempo real.

La cuantificación relativa de los niveles de expresión génica se llevó a cabo mediante la técnica de PCR en tiempo real (Real-Time PCR) según el método  $2^{-\Delta\Delta C_t}$  [111].

Como calibrador se utilizó un gen de expresión constitutiva que codifica por la Ser/Thr proteína fosfatasa 2 (PP2A; At1g13320) [39]. Un aspecto importante que se tuvo en cuenta para la aplicación del método de cuantificación  $2^{-\Delta\Delta C_t}$  es que las eficiencias de amplificación de los fragmentos del gen de interés y del gen de referencia sean similares y próximas a 2 [111].

Para el diseño de los oligonucleótidos para la cuantificación de blancos de miARNs se tuvieron en cuenta las siguientes consideraciones. Los miARNs en plantas regulan la expresión de sus blancos principalmente mediante el corte del ARNm. De esta manera para cuantificar el nivel de expresión de un gen que es blanco de un miARN es fundamental distinguir el ARNm completo del cortado. Dado que en la RT se sintetizan las hebras de ADNc a partir del oligo dT que hibrida en el extremo 3' de los ARN mensajeros, la secuencia que se ubica 5' del sitio complementario al miARN solamente será retrotranscripta en aquellos ARNm que no hayan sido cortados por el miARN. De esta manera si los oligonucleótidos utilizados en la PCR en tiempo real hibridan en esta región se amplificará específicamente un fragmento presente en el ADNc generado de ARNm intactos.

Las reacciones se realizaron en un equipo de qPCR “Mastercycler® ep realplex” (Eppendorf) en microtubos apropiados para esta técnica. Para determinar el perfil de amplificación durante la PCR se utilizó el colorante fluorescente “SYBR Green I Nucleic Acid Gel Stain” (Roche).

Las reacciones de amplificación se llevaron a cabo bajo las siguientes condiciones. Para un volumen final de reacción de 20  $\mu$ l se adicionó a cada tubo: 2  $\mu$ l de “10X PCR Buffer” (Invitrogen), 3mM MgCl<sub>2</sub>, 0,2mM de cada dNTP, 5  $\mu$ l de ADNc, 0,5 U de “Platinum Taq DNA Polymerase” (Invitrogen), 20 pmol de cada cebador, y 0,8  $\mu$ l de una dilución 1000X en

agua del “SYBR Green I Nucleic Acid Gel Stain (Roche)” original. El volumen se completó con agua Milli-Q.

Como control negativo se preparó un tubo de reacción para cada muestra en el cual se agregaron los mismos componentes en las mismas condiciones, pero en lugar de utilizar ADNc como molde se colocó 5  $\mu$ l del ARN sobrante luego del tratamiento con ADNasa, diluido en la misma proporción.

El termociclado de las mezclas de reacción se llevó a cabo según el programa que se muestra en la tabla 3.1.

Tabla 3.1: Programa de termociclado

Etap	Temperatura (°C)	Tiempo	Nº de ciclos
Desnaturalización	95	1 min	1
Ciclos	95	15 seg	40
	55	30 seg	
	72	40 seg	
	Lectura de fluorescencia		

Se incluyó una etapa final de determinación de la temperatura de fusión de los productos amplificados, lo que permite conocer la especificidad de la reacción. Además, la primera vez que se utilizaron los oligonucleótidos el producto de la reacción de amplificación se analizó en un gel de agarosa para confirmar la especificidad de la reacción.

Los oligonucleótidos utilizados en este estudio son que se detallan en la Tabla 3.2.

Tabla 3.2: Oligos utilizados para la RT-qPCR

Gene	Locus ID	Forward primer	Reverse Primer
PAA2	At5g21930	GTCCTCTTATCAGGGGACAGG	CATAGTTGCTTGTGCAAGACTCAG
MYB33	At5g06100	CTATGGAAACCGACATTCACCTG	CTTGGCTTCCAGAAGCAACATATCG
NZZ	At4g27330	TCGGGTCAGGTTATGATCGA	AGGGTTTCCTTCCATGTAGCTCC
PP2A	At1g13320	CCTGCGGTAATAACTGCATCT	CTTCACTTAGCTCCACCAAGCA
tMT2A	tobacco	TACCCAGATTTGAGCTACAACGAG	GCAGGAGATTCACCCATTTCCATA
tMT2B	tobacco	TACCCAGATTTGAGCTACAACGAA	AGGGGATTCACCCATTTCCATT

### 3.11. Mapeo del sitio de corte.

En la Figura 3.4 se muestra un esquema para la identificación de los productos de corte de miARNs. Inicialmente, se realizó una extracción 50mg de ARN total de plántulas de Col-0 y se realizó una purificación de ARN utilizando el kit comercial “PolyAtract®” de (Promega). El ARNm contenido en un volumen de 100  $\mu$ l de agua se precipitó mediante el agregado de

2  $\mu$ l 10 mg/ $\mu$ l glucógeno, 10  $\mu$ l de 3M Acetato Sodio pH 5.2 y 220  $\mu$ l de etanol absoluto, e incubación toda la noche a -20 °C. Luego se centrifugó 20 min a 4 °C a 20000 g. Se lavó con etanol 80%, y se centrifugó 5 min a 4 °C a 20000 g. Se removió el sobrenadante y se secó 10 min a 37 °C. Finalmente se resuspendió en 10  $\mu$ l agua.

A continuación se realizó la ligación del Oligo Adaptador ARN (5' CAGCUGGAGCAC-GAGGACACUGACAUGGACUGAAGGAGUAGAAA 3') mediante el siguiente protocolo. Se mezcló 250 ng de ARNm con 0,25  $\mu$ l de oligo ARN (100  $\mu$ M), y se incubó 5 min a 65 °C. Luego se incubó en hielo 2 min. Se agregó la mezcla de reacción: 1  $\mu$ l 10X Buffer Ligasa, 1  $\mu$ l 10mM ATP, 1  $\mu$ l BSA, 40 U RNaseOut (Invitrogen), 5 U T4 RNA ligasa (Invitrogen), agua para completar 10 $\mu$ l. Se incubó a 37 °C durante 1 hora, luego se centrifugó y se dejó en hielo. Se precipitó el ARNm mediante el protocolo antes descripto. Posteriormente se realizó una retrotranscripción y finalmente se eliminó el exceso de oligo ARN GeneRacer incubando 30 min a 37 °C con 0,04 g/ $\mu$ l de RNasa A.

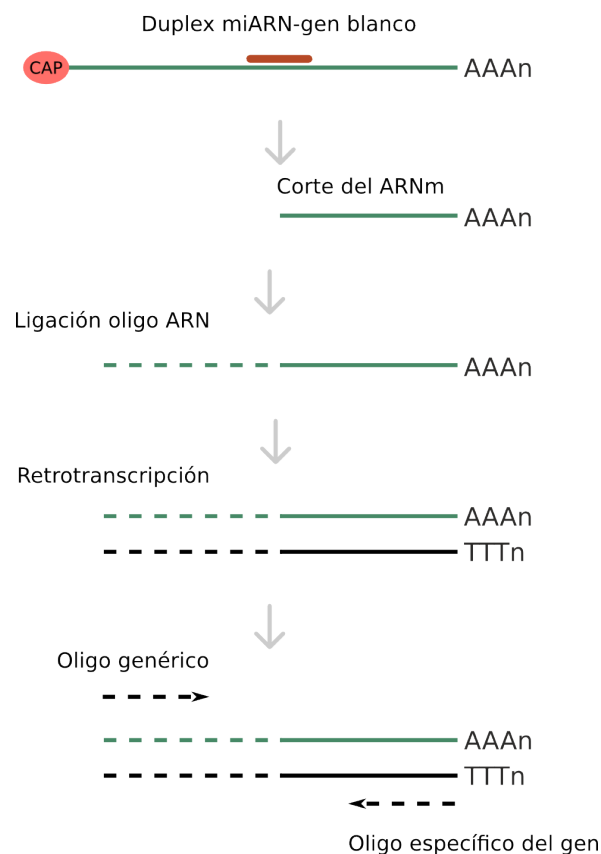


Figura 3.4: Diagrama de la estrategia utilizada para la identificación de los productos de corte de miARNs.



Para identificar los productos de corte específicos del gen en estudio se realizó una PCR con un oligo que hibrida sobre la secuencia del Oligo Adaptador ARN (oligogenérico) y otro oligo que hibrida sobre la secuencia del gen en una región 3' del sitio putativo de corte del miARN. Los productos de la PCR se clonaron y se secuenció al menos 15 clones independientes para determinar la secuencia del extremo 5'.

Los oligonucleótidos utilizados en este estudio son que se detallan en la Tabla 3.3.

Tabla 3.3: Oligonucleótidos utilizados para 5' RACE

Gen	Locus ID	5' RACE	5' RACE nested
General		CGACTGGAGCACGAGGACACTGA	GGACACTGACATGGACTGAAGGAGTA
PAA2	At5g21930	GACTTATGGAGCTGCAGAAGTAATG	CATAGTTGCTTGTGCAAGACTCAG
IAR3	At1g51760	ATCTTCTGATCCCATTAATGGTTGCATCTCG	CATATTCACGCTCGCTTGCCTTGTGATAACC
NZZ	At4g27330	CATTAAAGCTTCAAGGACAAATCAATGGTATTAGG	AGGGTTTCCTTCCATGTAGCTCC
MMG4.7	At5g43060	ATGGTAACAACCTTAGCAITTTTTC	CTTCGGTATCAATACCWCCAIT
UDP	At2g47650	AATGGGCCGACATGTTCTCC	CCTCGGTGATAGTCCATGGT
SVP	At2g22540	GCAACTTTCTTCATTCATC	TTTCATCTGCCTCAGCTCAC
loricrin-related	AT5g64550	ACCATGAGCTTTGCAGTAGT	CCTCAGCACTTCGTGTACAG
	At3g14110	CGGAAGGATCAGTCAGTCTC	CCCAGCTCGGTATAACAGTC
	At3g22110	GTTTCATCGCCAAAGGTAAC	CCAGGCGAATAAGACTAGAG
AVA-P2	At1g19910	CTCTAGACTGACCAGCTCGA	GGATGATACCAACAATGAGA

3.12. Protocolo de SPARE.

La técnica de SPARE (del inglés Specific Parallel Analysis of 5' RNA Ends) fue desarrollada con el objetivo de caracterizar el procesamiento de los precursores de miARNs de *A. thaliana*. El esquema de la técnica se muestra en la Figura 3.5 y el protocolo se preparó como se describe en Methods[124]

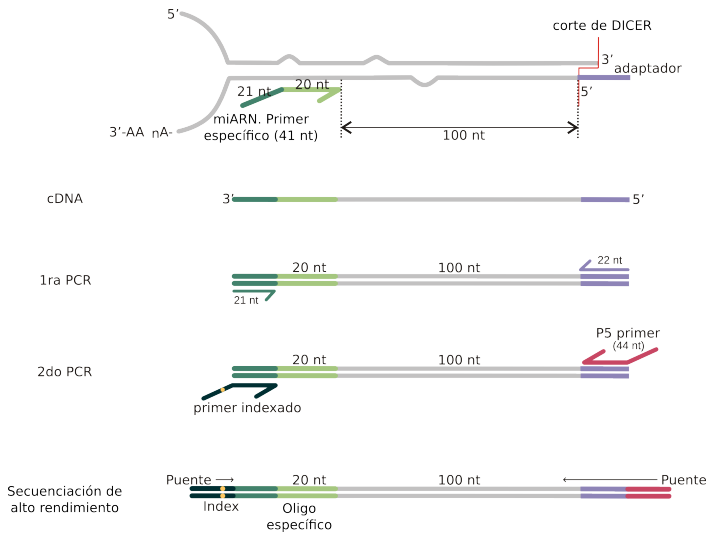


Figura 3.5: Esquema general para la construcción de bibliotecas de SPARE.

# Capítulo 1.

## Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco.

### 4.1. Introducción.

En plantas, los miARNs generalmente regulan más de un gen de una misma familia [67], así por ejemplo, miR319 regula a cinco factores de transcripción de la familia TCP, mientras que el miR172 hace lo mismo con seis factores de transcripción con dominio AP2. A su vez, los miARNs conservados en plantas se encuentran generalmente codificados como pequeñas familias de genes que codifican ARN pequeños de secuencias similares o idénticas (Figura 4.1). En el caso de miR319, hay tres genes diferentes que codifican para miARNs similares, en dos casos son idénticos (ath-miR319a, ath-miR319b), mientras que un tercero difiere en la base 20 (ath-miR319c, Figura 4.1).

Una de las posibles ventajas de tener familias génicas con múltiples miembros es que proporciona flexibilidad en la manera en la que cada uno de ellos es regulado [120, 121]. Estas diferencias pueden ser por variaciones en elementos regulatorios en los promotores o en las estructuras de los precursores de miARNs que llevan a eficiencias de procesamiento diferenciales [120].

Por otro lado, las diferencias en las secuencias de los miARNs maduros podrían causar que miARNs de una misma familia regulen diferentes conjuntos de genes. Esto se ha demostrado previamente para la superfamilia que incluye a los miARNs miR319 y miR159. Estos dos miARNs son muy similares en secuencia, al punto que varios autores los consideran miembros de una misma familia [67]. Sin embargo ha sido demostrado que regulan genes diferentes [107, 108, 127]. Así, mientras el miR319 puede reprimir la expresión de factores de transcripción de la familia TCP y MYB, diferencias de secuencia en 4 nucleótidos previenen la actividad del miR159 sobre los TCP [108].



Figura 4.1: **Conservación y divergencia de miARNs en distintas especies** Se muestra miembros de la familia del miR319 y del miR160 en distintas especies. En colores se muestra la conservación de las bases y el consenso (cons) se indica con asteriscos. Puede notarse que existe poca variabilidad de la secuencia del miARN maduro en distintas especies, pero aun así pueden observarse algunos cambios.

La identificación de genes blancos regulados por miARNs en plantas es muy importante para poder conocer el rol de los miARNs. En general esta identificación, se obtiene de diferentes estrategias computacionales donde tienen en cuenta la complementariedad con sus mensajeros blanco. Uno de los mayores desafíos es predecir los genes regulados por estos ARN pequeños con una baja frecuencia de predicciones falsas.

A partir de interacciones miARN-ARNm blanco validadas experimentalmente, se han determinado algunos parámetros empíricos para la identificación de genes blancos. Por un lado, el hecho de que determinados genes con un número de “mismatches” fuesen blancos de regulación por miARNs mientras que otros con igual número no, sugería que debían existir otros factores que afectarían la interacción de los miARN con los ARNm blanco. A partir de aquí es que algunas estrategias tienen en cuenta la posición relativa de los “mismatches” [127]. Es por esto que se han utilizado otros enfoques para abordar este problema, por ejemplo el requerimiento de que el sitio complementario al miARN en los ARNm estuviera conservados entre genes homólogos de Arabidopsis y arroz [149].

Más recientemente se desarrolló una tecnología, en este caso experimental, conocida como “degradoma”, que permite obtener el perfil global de los productos de degradación de los ARN mensajeros [1, 53]. Este enfoque permitió detectar in vivo y a nivel genómico la actividad de los miARNs al identificarse los productos de corte de los ARNm mediado por estos ARNs pequeños. Dicha información es muy útil para la identificación de nuevos posibles blancos.

En esta parte de la Tesis se profundizó esta cuestión y se desarrolló un método bioinformático para la predicción de genes regulados por miARNs conservados en plantas.

## **4.2. Resultados y Discusión.**

### **4.2.1. Selección de secuencias de miARNs presentes en distintas especies.**

Enfocamos nuestro análisis en 22 miARNs que están conservados en Angiospermas [38, 42] (Tabla 4.1). En general estos miARNs están codificados por pequeñas familias de hasta 32 miembros. En los genomas completos de Arabidopsis, poplar y arroz es común encontrar variaciones en la secuencia de los miARNs pertenecientes a una misma familia, especialmente en el primer nucleótido y los nucleótidos 20 y 21 [42].

Tabla 4.1: miARNs y sus genes blancos en plantas

miARN	Consenso (18 nt)	Targets conocidos <sup>(a,b)</sup>
miR156	GACAGAAGAGAGTGAGCA	factores de transcripción SPL
miR159	TTGGATTGAAGGGAGCTC	factores de transcripción MYB, <b>NOZZLE (NZL)</b>
miR160	GCCTGGCTCCCTGTATGC	factores de transcripción ARF
miR162	CGATAAACCTCTGCATCC	DCL1
miR164	GGAGAAGCAGGGCACGTG	factores de transcripción NAC
miR166	CGGACCAGGCTTCATTCC	factores de transcripción HDZip
miR167	GAAGCTGCCAGCATGATC	factores de transcripción ARF, <b>IAA-ALANINE RESISTANT 3 (IAR3)</b>
miR168	CGCTTGGTGCAGGTCGGG	AGO1
miR169	AGCCAAGGATGACTTGCC	factores de transcripción CCAAT-HAP2
miR171	TTGAGCCGTGCCAATATC	factores de transcripción GRAS
miR172	GAATCTTGATGATGCTGC	factores de transcripción AP2
miR319	TGGACTGAAGGGAGCTCC	factores de transcripción TCP
miR390	AGCTCAGGAGGGATAGCG	TAS RNA
miR393	CCAAAGGGATCGCATTGA	TIR1 proteins, F-BOX proteins
miR394	TGGCATTCTGTCCACCTC	proteínas F-BOX
miR395	TGAAGTGTTTGGGGGAAC	ATP-sulfurilasas, transportadores de sulfato
miR396	TCCACAGCTTTCTTGAAC	factores de transcripción GRF, <b>MMG4.7, FLUORESCENT IN BLUE LIGHT (FLU)</b>
miR397	CATTGAGTGCAGCGTTGA	Laccases
miR398	GTGTTCTCAGGTACACCC	Cu/Zn SODs, CytC oxidase protein subunit, Chaperona de cobre (CCS)
miR399	GCCAAAGGAGATTGCCC	Enzima E2 de conjugación de ubiquitina
miR408	TGCACTGCCTCTTCCCTG	Blue copper proteins, Laccases, <b>P-TYPE ATPase (PAA2), PAC1 (Proteasome component)</b>
miR827	TAGATGACCATCAGCAA	SPX proteins

a Los genes blancos fueron agrupados según sus funciones.

b Nuevos genes blancos validados experimentalmente en este estudio están indicados en negrita.

Sin embargo, observamos que la región entre la posición 2 y 19 está relativamente conservada y pudimos encontrar una secuencia consenso presente en la mayoría de los miembros de cada familia de miARNs en esas tres especies (Tabla 4.1). Curiosamente, las bases variables fuera de esta región conservada son propensas a tener “mismatches” con genes blancos conocidos, lo que indica que podría existir una correlación entre la interacción miARN-gen blanco y la conservación de la secuencia del miARN.

Nos enfocamos entonces en estas secuencias de 18 nt (entre las posiciones 2 y 19) en *Arabidopsis*, *poplar* y *arroz* que cuentan con sus genomas totalmente secuenciados. Vimos que si consideramos a estas especies podíamos encontrar una secuencia consenso de 18 nt correspondiente a cada una de las 22 familias de miARNs conservados y que a su vez exista como tal en todas las especies (Tabla 4.1). Dada la conservación de estas secuencias consenso de 18 nt en al menos estas tres especies distanciadas evolutivamente, y debido a que las bases de los extremos tienden a no interaccionar con las secuencias blanco, decidimos hacer nuestras búsquedas de blancos conservados evolutivamente con estos consensos.

#### 4.2.2. Selección de la biblioteca de transcriptos de plantas.

Para estudiar la regulación por miARNs de genes en distintas especies vegetales, decidimos recurrir a la información disponible públicamente en el “Gene Index Project”<sup>13</sup>. El proyecto es mantenido y administrado por la universidad de Harvard y contiene un catálogo completo de genes en una amplia gama de organismos incluyendo plantas. En este proyecto

<sup>13</sup><http://compbio.dfci.harvard.edu/tgi/>

las bibliotecas de EST (secuencias parciales de genes que se expresan) son ensambladas en “contigs” de tal manera de construir los ARNm que son expresados en las distintas especies. Estas bibliotecas cuentan por lo tanto con transcritos cuya expresión está validada empíricamente.

De este catalogo elegimos 41 especies de angiospermas (plantas con semillas) para hacer la búsqueda. Además, utilizamos ARNm completos para las especies de *A. thaliana* y *Oryza Sativa* que se encuentran caracterizados en detalle (para ver la lista completa de especies, ver Tabla 4.2).

### 4.2.3. Primera búsqueda general de potenciales genes blancos.

Utilizando las secuencias consenso de 18nt y permitiendo 3 “mismatches” (Tabla 4.1), la búsqueda de genes blancos en las 43 especies seleccionadas (Tabla 4.2) arrojó como resultado 38.597 genes (Figura 4.2, bin 1). En esta búsqueda las interacciones G-U y los “bulges” fueron considerados como “mismatches”. Todos los genes blancos de *A. thaliana* conocidos hasta ese momento fueron identificados usando esta estrategia con la excepción de CSD2, un gen blanco del miR398 que contiene 4 “mismatches” (tabla 4.2).

Teniendo en cuenta que la mayoría de los genes blancos arrojados presentan una escasa descripción del tipo genómica funcional, realizamos un BLASTx contra el proteoma de *A. thaliana*. El “locus ID” obtenido como “best hit” se utilizó como “tag” (etiqueta) para identificar al candidato en distintas especies (Figura 4.2). A pesar que esta estrategia no necesariamente identifica el gen ortólogo de Arabidopsis, sirve como propósito de clasificación de cada potencial gen blanco de miARN. Aunque la mayoría de los potenciales genes blancos pudieron ser fácilmente asignados con una etiqueta, algunos pocos casos, que incluye a los genes que representan ARNs no codificantes, fueron perdidos en este paso. Dado que ahora los genes de las distintas especies tienen una etiqueta común, podemos saber para cada gen, en cuales especies detectamos una interacción con un miARN.

Este enfoque permite la selección de los mejores candidatos basándose en la presencia de los genes blancos en un número mínimo de especies. Utilizando 4 especies como el mínimo de especies requeridas (ya que tiene una buena especificidad), dio como resultado 3.781 genes que corresponden a 533 etiquetas diferentes (Figura 4.2, bin 2). Esto significa que hay alrededor de 533 genes diferentes (etiquetas) dan interacción positiva en un promedio de 7 especies cada uno (es decir, detectamos 3.781 genes en el total de las bibliotecas analizadas).

Tabla 4.2: Especies y base de datos utilizadas para la búsquedas de genes blancos de miARNs conservados

<b>Especie</b>	<b>Base de datos</b>
Allium cepa	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Aquilegia	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Arabidopsis thaliana	<a href="http://arabidopsis.org/">http://arabidopsis.org/</a>
Beta vulgaris	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Brassica napus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Capsicum annuum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Citrus clementina	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Citrus sinensis	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Coffea canephora	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Euphorbia esula	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Festuca arundinacea	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Glycine max	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Gossypium	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Gossypium raimondii	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Haseolus vulgaris	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Helianthus annuus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Hordeum vulgare	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Ipomoea nil	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lactuca sativa	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lactuca serriola	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Lotus japonicus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Malus x domestica	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Medicago truncatula	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Mesembryanthemum crystallinum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Nicotiana benthamiana	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Nicotiana tabacum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Oryza sativa	<a href="http://www.jcvi.org/">http://www.jcvi.org/</a>
Panicum virgatum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Petunia hybrida	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Phaseolus coccineus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Populus	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Prunus persica	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Saccharum officinarum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Secale cereale	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Solanum lycopersicum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Solanum tuberosum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Sorghum bicolor	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Theobroma cacao	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
triphysaria	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Triphysaria versicolor	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Triticum aestivum	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Vitis vinifera	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>
Zea mays	<a href="http://compbio.dfci.harvard.edu/tgi/">http://compbio.dfci.harvard.edu/tgi/</a>

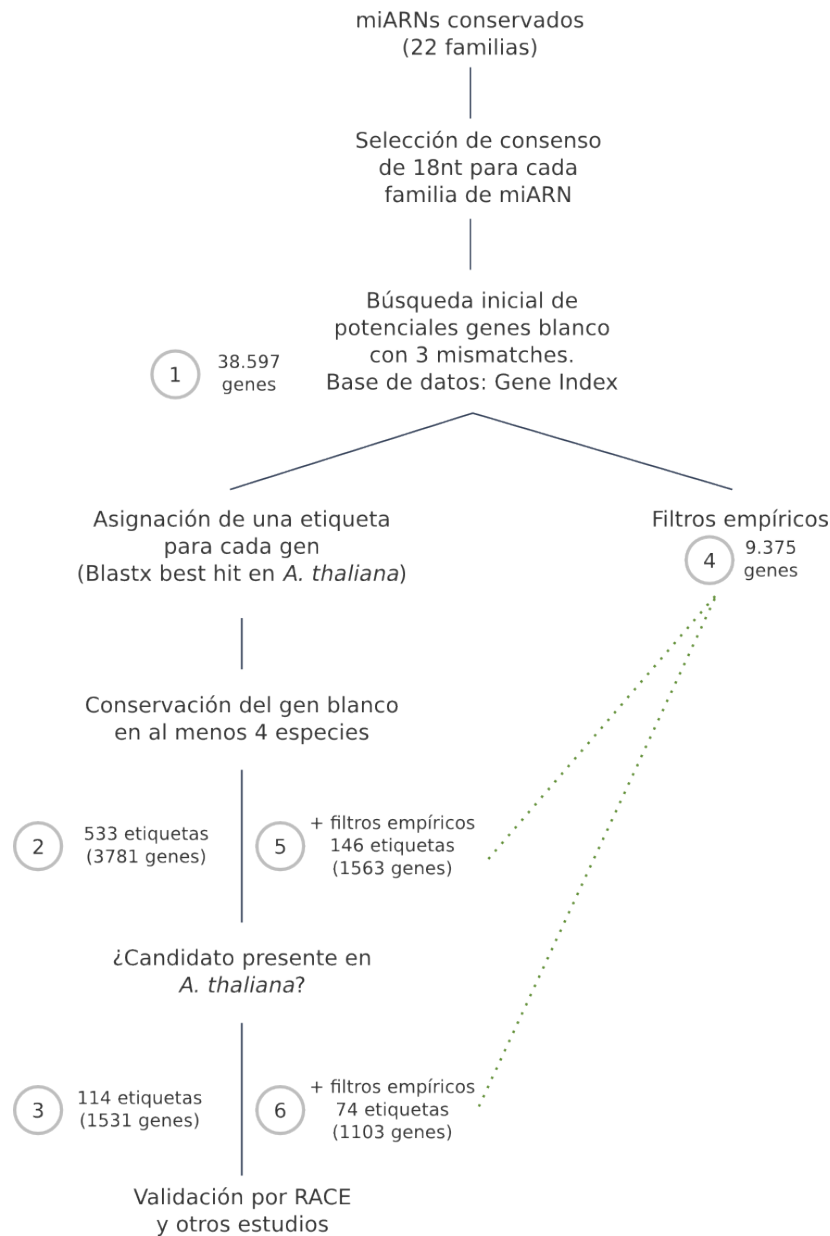


Figura 4.2: **Esquema de la estrategia para la identificación de nuevos genes blancos.** El número de genes blancos está identificado en cada paso. Luego de aplicar el análisis de conservación, todos los genes que tienen el mismo hit en Arabidopsis, fueron considerados como un solo gen blanco. El lado derecho muestra la búsqueda hecha con filtros empíricos: bin 5 y 6 incluyen genes blancos seleccionados con ambos filtros, empíricos y de conservación. Mientras que el bin 2 y 3 muestra los potenciales genes blancos seleccionados sólo con el filtro de conservación.



La búsqueda también se puede hacer en combinación con filtros empíricos de interacción el par miARN-gen blanco que tienen en cuenta la energía de interacción (mínima energía libre de hibridación de al menos 72 % del apareamiento perfecto) y la posición de los “mismatches” (no más de un “mismatch” entre la posición 1 y 11) (para más detalles ver Materiales y métodos sección 3.1.2). De los 38.597 candidatos iniciales, 9.375 pasan estos filtros (Figura 4.2, bin 4). Combinando filtros de energía y filtro de conservación evolutiva, la búsqueda arrojó como resultado 563 candidatos correspondientes a 146 etiquetas (Figura 4.2, bin 5).

#### **4.2.4. Parámetros empíricos y de conservación evolutiva pueden actuar de manera sinérgica para identificar genes blancos regulados por miARNs.**

Potenciales genes blancos de miARNs fueron clasificados de acuerdo al mínimo número de especie en donde fueron detectados (Figura 2A-E). Como control para cada miARN generamos 10 secuencias “scramble” (al azar), dividiendo las secuencias originales de los miARN consensos en di-nucleótidos y luego generando nuevas secuencias al azar conservando la composición de los di-nucleótidos. Estas secuencias al azar fueron utilizadas para realizar búsqueda de genes blancos del mismo modo que lo hicimos para las secuencias originales. Luego, la relación señal/ruido fue calculada como el cociente entre el número de genes blancos para los miARNs y el número promedio obtenido de las secuencias al azar. Detectamos una relación señal/ruido de 1,2 para todos los miARNs juntos, utilizando como condiciones de interacción tres “mismatches” para los miARN consenso.

A continuación, empezamos a requerir que haya conservación de la interacción en distintas especies, es decir, que el mismo par miARN/gen blanco (“tag”), sea detectado en un número mínimo de especies. A partir de aquí, observamos que la relación señal/ruido incrementa con el número de especie en donde los genes blancos fueron detectados (Figura 4.3 A, recuadro). Los datos para todos los miARNs y sus potenciales genes blancos conservados en al menos 4 especies están incluidos en la tabla 4.3.

Luego estudiamos la selección de candidatos teniendo en cuenta los filtros empíricos. Para esto aplicamos una versión modificada de los filtros descritos anteriormente, y requiriendo (i) una energía mínima de hibridación (MFE) de al menos 72 % del apareamiento perfecto que tendría cada secuencia consenso y (ii) que sólo un “mismatch” pudiera estar presente entre la posición 1 y la 11 de la secuencia consenso (2-12 del miARN). De la búsqueda inicial

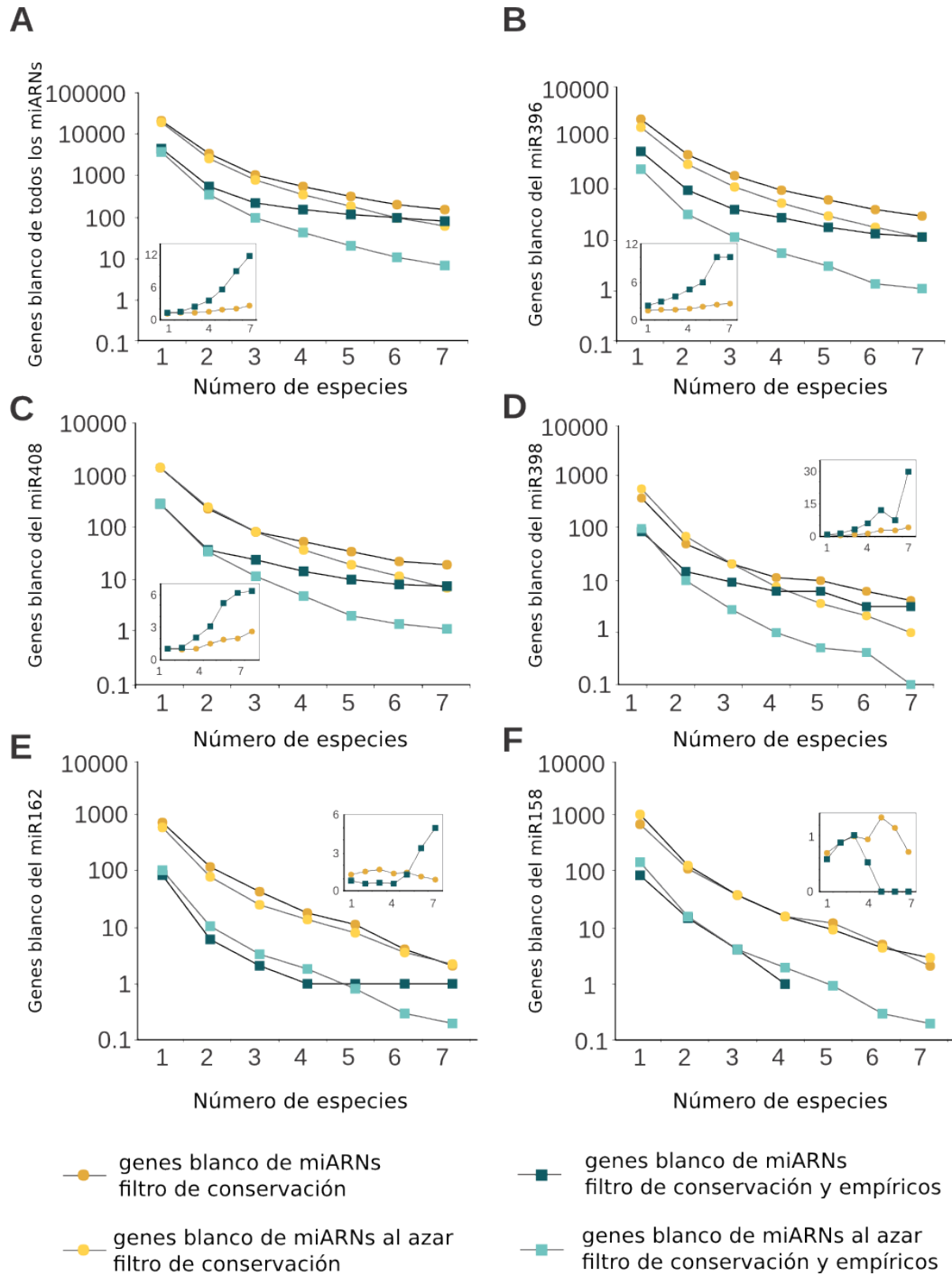


Figura 4.3: **Conservación de potenciales genes blancos en distintas especies.** Todos los miARNs (A), miR396 (B), miR408 (C), miR398 (D), miR162 (E), miR158 (F). Puntos naranja representan los genes blancos de miARNs usando filtro evolutivo. Puntos amarillos representan los genes blancos de las secuencias al azar usando filtro evolutivo. El cuadrado azul muestra los genes blancos de miARNs luego de aplicar filtros empíricos y evolutivos, mientras que el cuadrado celeste representa los genes blancos de las secuencias al azar en las mismas condiciones. Los recuadros muestran la relación señal/ruido.

9.375 genes pasaron estos filtros conteniendo el 97 % de los genes validados anteriormente de *Arabidopsis*. (Figura 4.2, bin 4).

Al aplicar solamente este filtro empírico, dio como resultado una relación señal/ruido de 1,7 cuando agrupamos todos los miARNs juntos (Figura 4.3 A). Observamos que aplicar simultáneamente los filtros empíricos y de conservación aumentó significativamente la relación señal/ruido para todos los miARNs juntos (Figura 4.3 A recuadro) y también de cada miARN individualmente (Figura 4.3 B-E, recuadros y tabla 4.3). En varios casos, esta relación es de hasta 10 veces cuando se requiere de que el gen blanco esté presente en más de 5 especies y que pase los filtros empíricos (Figure 4.3 A–D). Este efecto sinérgico indica que el filtro de conservación evolutiva y los parámetros empíricos pueden estar seleccionando aspectos diferentes de la interacción miARN-gen blanco.

Observamos que el número de genes blancos candidato y la relación señal/ruido es variable entre los distintos miARNs. El miR396 tiene la mayor cantidad de potenciales genes blancos, 92 de ellos presentes en al menos 4 especies y 26 de ellos pasan además los filtros empíricos (Tabla 4.3 y Figura 4.3 B). El miR408 y el miR398 también tienen un número alto de potenciales genes blancos y buenas relaciones de señal/ruido (Figura 4.3 C-D).

En contraste, ciertos miARNs como el miR162, miR168 y miR399 tienen un solo potencial gen blanco conservado en al menos 4 especies de acuerdo con nuestra búsqueda (Tabla 4.3 y Figura 4.3 E). Al menos en el caso del miR162 y del miR168 este resultado podría estar reflejando su rol específico en la regulación por retroalimentación de la biogénesis del miARN, ya que controlan los niveles de expresión DCL1 y AGO1 respectivamente [145, 155].

Como control adicional para nuestra estrategia, realizamos la búsqueda de genes blancos del miR158 y miR173, que son miARNs presentes solamente en *A. thaliana* y especies bien cercanas (17). Como era esperado, estos miARNs no generaron más candidatos que sus versiones al azar cuando los analizamos en las bibliotecas correspondientes a angiospermas en general (Tabla 4.3 y Figura 4.3 F).

Luego chequeamos si los pares miARN-gen blanco altamente conservados tenían una interacción más fuerte que los que están presentes en pocas especies. Para esto calculamos la energía mínima de hibridación para cada interacción detectada en nuestro trabajo. Observamos que los pares miARN-gen blanco presentes en muchas especies tienden a tener energía de interacción mayores que los que están presentes en menos especies (Figure 4.4 A). De todos modos, la correlación no fue notoria y algunas interacciones miARN-gen blanco

tuvieron una baja energía de hibridación (Figure 4.4 A). Estos resultados muestran que una alta conservación podría no ser necesariamente equivalente a una fuerte interacción, la misma podría proporcionar una explicación para los efectos sinérgicos causados por los filtros de evolución y empíricos sobre la relación señal/ruido.

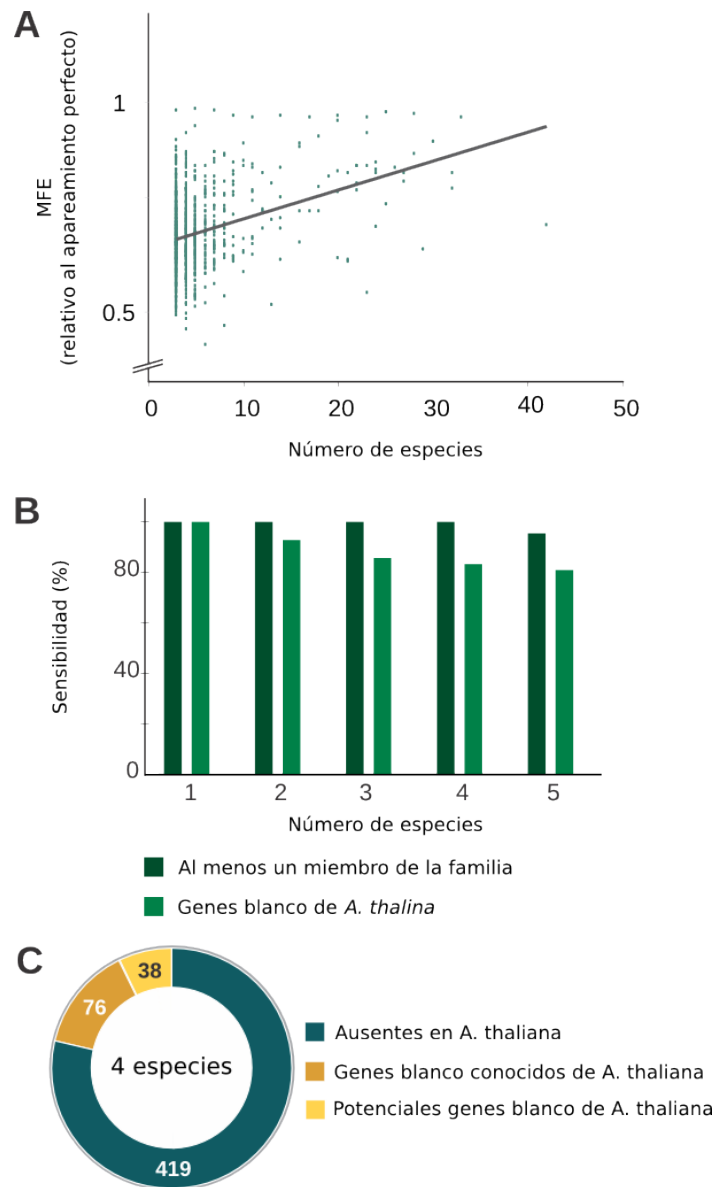


Figura 4.4: **Selección de genes blancos por conservación evolutiva de la secuencia.** (A) Relación entre MFE y el número de especies en donde cada gen blanco fue detectado. (B) Sensibilidad de la estrategia, analizado de dos modos distinto. Verde claro: evaluando la presencia de genes validados en *Arabidopsis* y en verde oscuro teniendo en cuenta la presencia de por lo menos un gen blanco de cada familia regulada por miARNs. (C) Clasificación de los potenciales genes blancos presentes en al menos 4 especies.

Tabla 4.3: Detección de genes blancos de miARNs utilizando diferentes filtros

	Sin filtros			Filtros empíricos			Conservación 4 especies			Todos los filtros		
	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio	miARN	scramble	ratio
<b>miR156</b>	3915	3994.4 ± 149.9	1.0	890	704.7 ± 45.2	1.3	34	39.7 ± 3.1	0.9	10	5.4 ± 1.1	1.9
<b>miR159</b>	1663	1283.7 ± 47.8	1.3	472	254.9 ± 21.9	1.9	20	10.1 ± 1.1	2.0	6	1.5 ± 0.5	4.0
<b>miR160</b>	793	695.6 ± 30.5	1.1	277	157.5 ± 28.8	1.8	5	4.4 ± 0.9	1.1	4	0.5 ± 0.3	8.0
<b>miR162</b>	1191	930.2 ± 139.5	1.3	108	164.7 ± 24.1	0.7	18	13.5 ± 3.5	1.3	1	1.8 ± 0.5	0.6
<b>miR164</b>	2486	1480.2 ± 60.4	1.7	678	333.2 ± 32.2	2.0	39	12.4 ± 1.9	3.1	12	1.5 ± 0.5	8.0
<b>miR166</b>	879	815.5 ± 45.0	1.1	231	129 ± 14.5	1.8	16	10.6 ± 1.4	1.5	6	0.9 ± 0.4	6.7
<b>miR167</b>	1777	1364.2 ± 146.6	1.3	478	214.8 ± 27.5	2.2	22	20.2 ± 3.6	1.1	4	1.8 ± 0.5	2.2
<b>miR168</b>	962	797.5 ± 48.5	1.2	209	185 ± 14.2	1.1	6	4.4 ± 0.8	1.4	1	1.1 ± 0.5	0.9
<b>miR169</b>	1540	1047.2 ± 69.7	1.5	464	181.4 ± 15.6	2.6	26	11.1 ± 2.1	2.3	10	1.2 ± 0.2	8.3
<b>miR171</b>	884	723.4 ± 32.1	1.2	202	113.8 ± 13.4	1.8	7	6.6 ± 1.4	1.1	2	0.7 ± 0.3	2.9
<b>miR172</b>	3007	1693.7 ± 124.7	1.8	540	288.1 ± 40.3	1.9	34	17.7 ± 1.7	1.9	5	2.2 ± 0.6	2.3
<b>miR319</b>	1363	1274.2 ± 113.6	1.1	324	249.2 ± 22.3	1.3	18	15 ± 2.8	1.2	7	1.8 ± 0.5	3.9
<b>miR390</b>	873	814.4 ± 64.3	1.1	335	173 ± 22.5	1.9	8	4.7 ± 1.2	1.7	3	0.7 ± 0.5	4.3
<b>miR393</b>	986	844.6 ± 58.7	1.2	276	124.6 ± 11.1	2.2	14	7.1 ± 1.2	2.0	5	0.5 ± 0.2	10.0
<b>miR394</b>	1569	1531.4 ± 57.5	1.0	188	237.1 ± 25.0	0.8	26	21.4 ± 2.2	1.2	3	2.9 ± 0.5	1.0
<b>miR395</b>	1472	1226.7 ± 66.7	1.2	426	217.6 ± 16.5	2.0	11	8.8 ± 1.3	1.3	6	1.3 ± 0.3	4.6
<b>miR396</b>	4641	2979.3 ± 246.6	1.6	1246	390.5 ± 38.8	3.2	92	51.4 ± 5.9	1.8	26	5.4 ± 1.0	4.8
<b>miR397</b>	1426	1050.9 ± 27.9	1.4	368	236.5 ± 23.5	1.6	26	9.7 ± 0.8	2.7	10	1.6 ± 0.3	6.3
<b>miR398</b>	935	834 ± 34.5	1.1	376	144 ± 18.1	2.6	11	7.5 ± 1.6	1.5	6	1 ± 0.3	6.0
<b>miR399</b>	1192	1137.6 ± 72.0	1.0	275	207.8 ± 24.9	1.3	5	13.6 ± 1.7	0.4	1	1.5 ± 0.7	0.7
<b>miR408</b>	2782	2502.9 ± 103.6	1.1	695	468.7 ± 50.8	1.5	51	35.1 ± 3.0	1.5	14	4.6 ± 0.8	3.0
<b>miR827</b>	2261	2000.1 ± 119.8	1.1	317	297.1 ± 45.0	1.1	44	23.4 ± 3.9	1.9	4	2.3 ± 0.8	1.7
<b>Total</b>	38597	31021.7 ± 1859.8	1.2	9375	5473.2 ± 576.3	1.7	533	348.4 ± 47.0	1.5	146	42.2 ± 11.3	3.5
<b>Control</b>					±							
<b>miR158</b>	1364	1462.8 ± 69.1	0.9	170	208.7 ± 15.8	0.8	15	16 ± 1.7	0.9	1	1.9 ± 0.4	0.5
<b>miR173</b>	1386	1232.1 ± 101.7	1.1	243	215.6 ± 23.4	1.1	11	12 ± 2.4	0.9	1	1.5 ± 0.4	0.7

a Sin filtros, búsqueda inicial utilizando los miARN consenso de 18nt y 3 “mismatches”.

b Filtros empíricos, energía de al menos 72 % del apareamiento perfecto y 1 “mismatch” en la posición 2-12 del par miARN-gen blanco.

c Conservación del ID “tag” en al menos cuatro especies.

d Todos los filtros, combinación de los filtros empíricos y de conservación en al menos cuatro especies.

e miARN, genes blancos para cada miARN específico.

f scramble, promedio de los genes blancos de 10 versiones al azar de cada miARN ± error estándar.

#### 4.2.5. Identificación de nuevos genes blancos en *A. thaliana* por conservación de la secuencia del gen blanco.

Para encontrar nuevos genes blancos nos enfocamos en los genes potenciales que fueron seleccionados por nuestra estrategia utilizando solamente conservación evolutiva, debido a que los parámetros empíricos ya fueron utilizados extensamente en trabajos anteriores [5, 66, 127]. En primer lugar, analizamos la detección de genes blancos validados previamente en *A. thaliana* [basado en [48]]. Usando nuestra estrategia encontramos que el 84 % de ellos estaban presentes en al menos 4 especies (Figura 4.4 B). Consideramos esto como un buen resultado ya que puede ser que no todos los genes blancos de *Arabidopsis* estén conservados evolutivamente.

Generalmente los miARNs en plantas regulan genes que codifican para proteínas de la misma familia (Tabla 4.1), es por esto que evaluamos si por lo menos un miembro de cada familia era detectado en nuestro enfoque. Encontramos genes blancos pertenecientes a casi todas las familias de genes codificantes para proteínas presentes en cuatro especies (Figura 4.4 B), con la excepción de TAS3, que es regulado por el miR390, al ser un ARN no codificante no es detectado por Blastx.

Para encontrar nuevos genes blancos regulados por miARNs, nos enfocamos únicamente en los potenciales genes blancos conservados en 4 especies, donde una de ellas es *A. thaliana* (Figura 1, bin 1). Genes blancos de miARNs que no están presentes en *A. thaliana* podrían incluir genes que perdieron su regulación durante la evolución o genes que hayan adquirido control por un miARN conservado más reciente en otras especies. La conservación en cuatro especies fue elegida como filtro evolutivo porque provee buena sensibilidad para genes blancos conocidos.

Identificamos 114 potenciales genes que satisfacen este criterio. Donde 76 de ellos son genes validados anteriormente o genes muy relacionados (Figura 4.4 C). Curiosamente encontramos 38 genes que no tienen relación con genes blancos conocidos de miARNs y decidimos estudiar este grupo con mayor detalle. Nos enfocamos primero en los genes que estaban presentes en un gran número de especies para tener mejor especificidad (Figura 4.3).

Un potencial gen blanco del MiR408 era At5g21930 que codifica para P-TYPE ATPase OF ARABIDOPSIS 2 (PAA2) y estaba presente en 22 especies distintas incluido monocotiledóneas y dicotiledóneas. MiR408 es inusual debido a que tiene un 5'-A, sin embargo >30 % de las secuencias maduras del miR408 corresponden a una variante corrida 1 nt que empieza con 5'-U [95] (Figura 4.5 A). La validación experimental fue realizada por la técnica "5'

RACE PCR” (ver Materiales y Metodos sección 3.11) por Valeria Crosa, y la misma reveló fragmentos de ARNm compatible con este último sitio de corte (Figura 4.5 A). PAA2 es necesaria para el transporte de iones de cobre a plastocianina [102], y su regulación por el miR408 está relacionada con el rol de este miARN en la homeostasis de cobre [157].

Otro potencial candidato del miR408 era At3g22110 que codifica para PROTEASOME ALPHA SUBUNIT C1 (PAC1) y estaba presente en 20 especies. Por medio de 5’ RACE PCR se observó que este gen es gen blanco del miR408 (Figura 4.5 A). Curiosamente la interacción del par miARN-gen blanco tiene 3 “mismatches” en la región 5’, y se hubiera perdido como potencial gen blanco si se aplicaban solamente los filtros empíricos.

Luego estudiamos los genes blancos del miR396, donde los genes SVP y SUI1 estaban presentes en 29 y 19 especies respectivamente. Pero en ambos casos fallamos al obtener producto de la PCR utilizando 5’ RACE PCR modificada. La falta de regulación de este gen por el miR396 podría estar relacionado a la débil energía de hibridación del par miARN-gen blanco, aunque no podemos descartar que el miR396 esté controlando su traducción. Más recientemente, se ha propuesto que miR396 regularia a SVP disminuyendo la vida media del transcripto, en vez de guiar el corte directamente [159].

Otros dos potenciales genes blancos del miR396 eran At5g43060 y At3g14110 que codifican para la proteasa MMG4.7 y FLUORESCENT IN BLUE LIGHT (FLU), respectivamente. Y en ambos casos se pudo detectar el corte (Figura 4.5 C y D).

En contraste con el miR408 y miR396, donde tienen varios potenciales genes blanco, obtuvimos un solo potencial gen blanco para el miR159, un factor de transcripción MYB que regula desarrollo del estambre y polen [98]. El otro potencial gen blanco era At4g27330, conocido como NOZZLE/SPOROXYLETLESS. Este factor de transcripción, que participa en desarrollo del estambre y óvulo [126, 161], fue también validado por 5’ RACE PCR (Figura 4.5 E). Es interesante notar que al menos las funciones de NOZZLE y PAA2 pueden estar directamente relacionadas con el rol de genes blanco, ya descritos anteriormente, del miR159 y miR408 respectivamente.

PAA2, FLU y NOZZLE fueron detectados en mono y dicotiledóneas mientras que PAC1 y MMG4.7 fueron detectadas solamente en dicotiledóneas (Figura 4.5 A-E). Las posiciones del sitio de unión del miARN-gen blanco están altamente conservadas y muchas de las posiciones variables corresponden a “mismatches” con el miARN o variaciones del tipo G-C/G-U. Además este método no requiere que el sitio del gen blanco esté conservada, sino más bien que haya una interacción predicha con el miARN en distintas especies. De esta

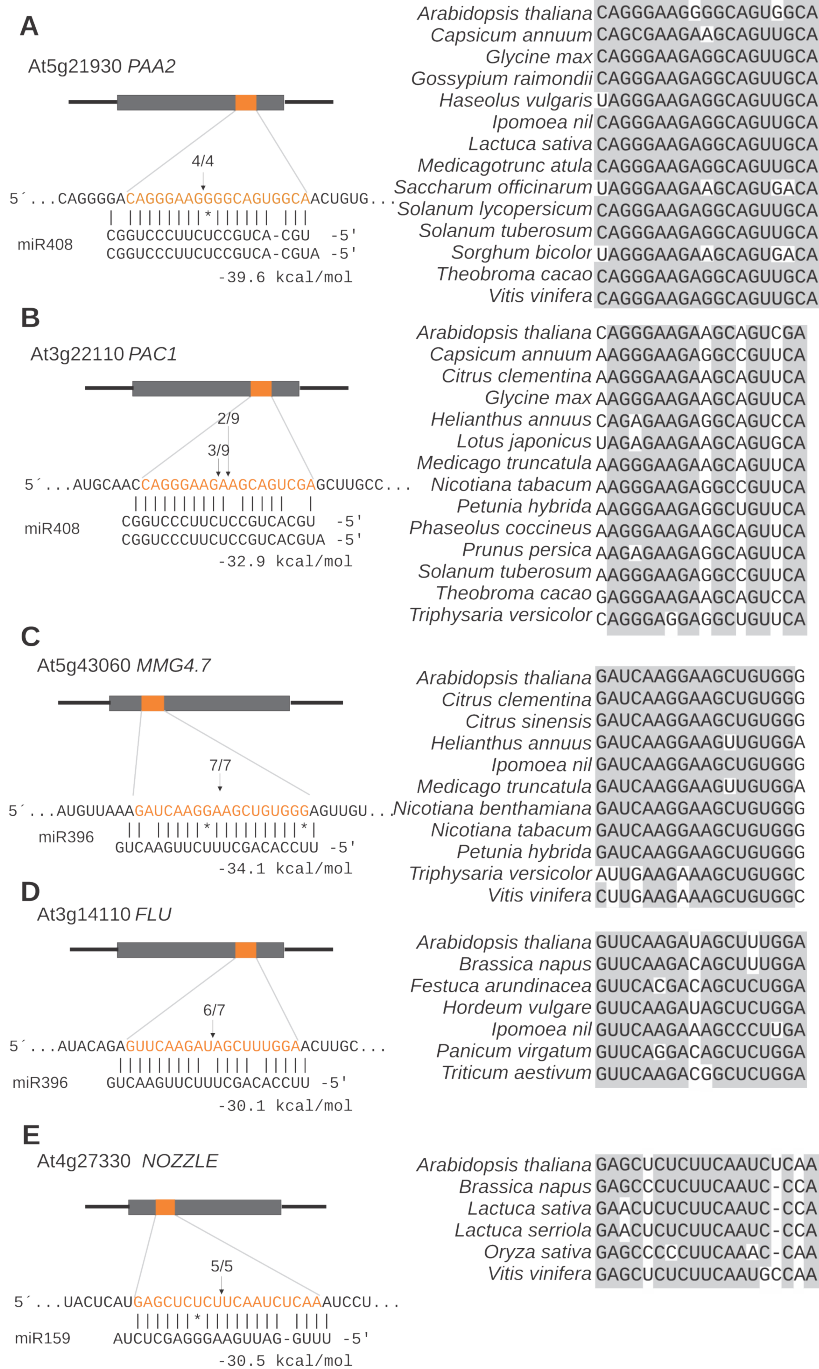


Figura 4.5: **Nuevos genes blancos validados en *A. thaliana*.** El alineamiento entre el miARN y los nuevos genes blancos identificados se muestran en la izquierda. La conservación evolutiva de la secuencia del sitio reconocido por el miARN en las especies seleccionadas se muestra a la derecha. La figura muestra las interacciones del miR408 con PAA2 (A), miR408 con PAC1 (B), miR396 con MMG4.7 (C), miR396 con FLU (D), miR159 con NOZZLE (E). Las flechas marcan el sitio de corte determinado por 5'RACE-PCR y los números indican la frecuencia de clonado de cada fragmento (la validación experimental fue realizada por Valeria Crosa).



manera el sitio de NOZZLE, donde la secuencia cambia en diferentes especies (Figura 4.5 E), pudo ser detectado por este enfoque.

#### **4.2.6. Identificación de nuevos genes blancos permitiendo interacciones G-U.**

Los genes blancos identificados utilizando la estrategia descrita anteriormente, tienen varios “mismatches” y “bulges” con sus miARNs, lo que puede ayudar a explicar por qué no fueron detectados en trabajos anteriores. También notamos que muchas de estas nuevas interacciones miARN-gen blanco contenían posiciones que variaban alternadamente entre G-C y G-U en distintas especies. Como consideramos G-U como “mismatch” en nuestra búsqueda inicial, decidimos realizar nuevamente la búsqueda con los miARNs consenso de 18nt pero permitiendo ahora 4 “mismatches”, donde al menos uno de ellos tiene que ser del tipo G-U. Esta búsqueda permitiría interacciones miARN-gen blanco con sólo 14 bases apareadas perfectamente.

Para compensar el uso de estos parámetros relajados en términos de “mismatches”, requerimos que el gen blanco aparezca en al menos 10 especies distintas para aumentar la especificidad (Figura 4.6 A). Encontramos 125 potenciales genes blancos en *A. thaliana* teniendo en cuenta este criterio (Figura 4.6 A) y 34 de ellos no aparecían en las búsquedas anteriores. El gen blanco CSD2 regulado por el miR398, que no apareció anteriormente, fue detectado con estos parámetros.

Luego examinamos el último grupo de potenciales genes regulados por miARNs que estaban realizando funciones auxiliares a los genes blancos ya descritos para cada miARN. Y encontramos que el miR167 que regula factores de respuesta a auxina (ARFs), también regulaba potencialmente a un gen denominado IAA-ALANINE RESISTANT 3 (IAR3) (Figura 4.6 B y C), que está involucrado en el control de niveles libre de auxina [41, 114].

IAR3 en *Arabidopsis* tiene 3 “mismatches” con respecto al miR167, pero en la posición 12 de la interacción miARN-gen blanco, hay una interacción G-U en varias especies (Figura 4.6 B y C). La técnica de 5' RACE PCR confirmó que el gen realmente era gen blanco del miR167 (Figura 4.6 C).

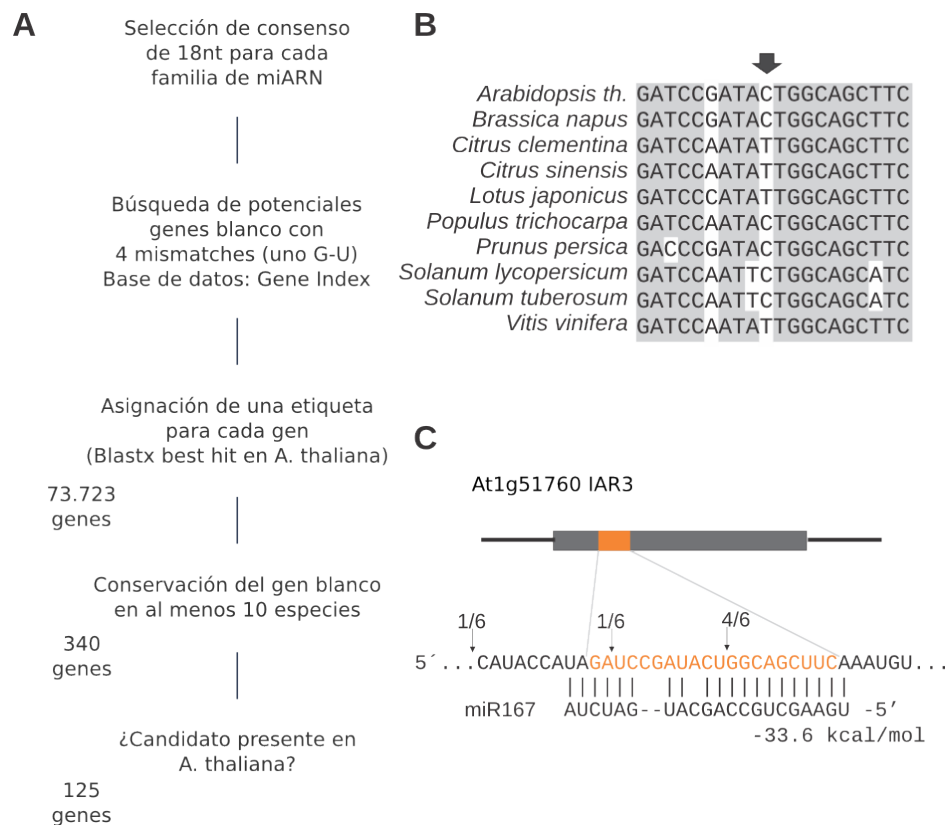


Figura 4.6: **Identificación de un nuevo gen blanco de miARN, relajando los parámetros de interacción pero incrementando el parámetro de conservación evolutiva.** (A) Esquema de la estrategia modificada para identificar genes blancos de miARNs. (B) Conservación del sitio blanco reconocido por el miARN en distintas especies. La flecha indica una variación de G-C o G-U con el miARN dependiendo de la especie. (C) Alineamiento en *A. thaliana* del gen blanco IAR3 con el miR167. La flecha indica la posición del corte indicada por 5'RACE-PCR y el número indica la frecuencia de clonado de cada fragmento.

#### 4.2.7. Identificación de genes blancos específicos de *Solanaceae*.

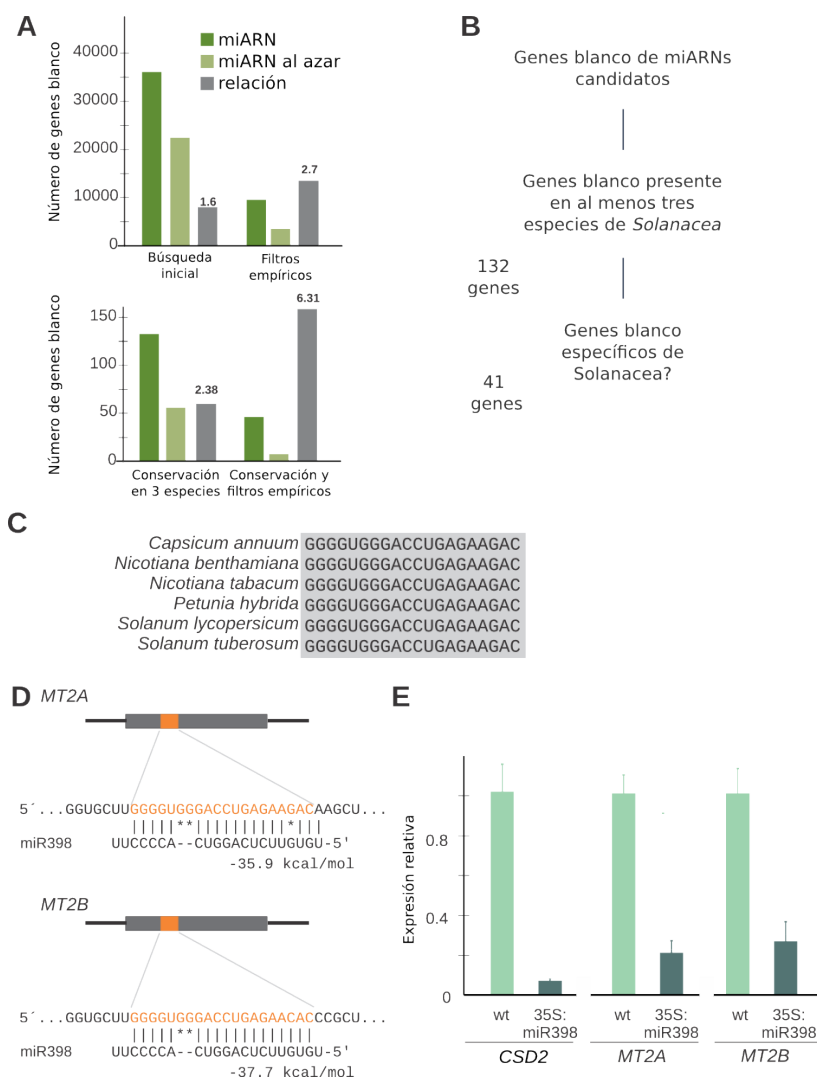
Pensamos que la estrategia mostrada también se puede utilizar para encontrar genes blancos presentes específicamente en un grupo de especies relacionadas. Por lo tanto, intentamos demostrar esto, encontrando potenciales genes blancos específicos de la familia de *Solanaceae*.

Elegimos esta familia en particular, ya que 6 especies estaban bien representadas en la biblioteca utilizada. La relación señal/ruido entre los genes blancos y las secuencias al azar era más de 2 cuando el filtro empírico o de conservación (en al menos 3 de las 6 especies *Solanaceae*) fueron aplicados (Figura 4.7 A). Curiosamente, al aplicar ambos filtros dio como resultado una relación señal/ruido por encima de 6 (Figura 4.7 A), confirmando nuestros previos hallazgos de que ambos filtros mejoran la detección de genes blancos de miARNs.

Encontramos 132 potenciales genes blancos presentes en al menos 3 especies *Solanaceae*. De este grupo, 41 genes no fueron detectados en otras especies (Figura 6B). El gen blanco más común fue la metalotioneína MT2A, presente en las 6 *Solanaceae*, como potencial gen blanco del miR398, mientras que MT2B, homólogo de este gen, fue detectado en 5 especies (Figura 4.7 B-D).

Luego, aprovechamos las plantas transgénicas de tabaco que contienen un transgén 35S:miR398 (A.F. Lodeyro, N. Carrillo y J.F. Palatnik resultados no publicados) y chequeamos la expresión de estos genes. Encontramos que CSD2, un gen blanco conservado del miR398, disminuía su expresión >10 veces en las plantas transgénicas 35S:miR398 comparadas con la planta salvaje (Figura 4.7 E). Curiosamente, observamos que tanto MT2A como MT2B disminuyeron sus niveles de transcripción >5 veces en estas plantas (Figura 4.7 E).

Estos resultados concuerdan con la regulación de MT2A y MT2B por el miR398, aunque no necesariamente demuestra una interacción directa. Además, estos resultados demuestran que los genes blancos presentes en un grupo específico de especies pueden ser encontrados utilizando esta estrategia. Estos experimentos en tabaco los realizó la Dra. Anabella F. Lodeyro, que pertenece al grupo del Dr. Nestor Carillo del IBR.



#### **4.2.8. comTAR: una herramienta para la predicción de genes blancos regulados por miARNs en plantas.**

A partir de la estrategia descrita en el capítulo anterior, que fue utilizada para encontrar y validar experimentalmente genes blancos regulados por miARNs en *A. thaliana*, desarrollamos una herramienta web denominada comTAR (Conserved plant miRNA target prediction tool) (<http://rnabiology.ibr-conicet.gov.ar/comtar>) [36]. La misma se puede utilizar para predecir potenciales genes blancos regulados por miARNs en plantas y está basada en la conservación evolutiva del par miARN-gen blanco con un número relajado de “mismatches”. ComTAR permite distintas opciones/parámetros de búsqueda que pueden ser modificados por el usuario:

- Filtro de “mismatch”: Solamente un “mismatch” está permitido entre la posición 1 y la 11 de la secuencia del miARN consenso (Sí/No).
- Corte por energía de hibridación: Se define que un gen blanco es predicho si la mínima energía de hibridación está por debajo del corte elegido.
- El número mínimo de especies donde una misma etiqueta está presente para un miARN particular.

Además, con la herramienta comTAR se pueden realizar diferentes búsquedas, pudiendo elegir entre distintas opciones que más se ajustan a la necesidad del usuario. Estas opciones de búsqueda se detallan a continuación.

##### **Realizar la búsqueda de potenciales genes blancos a partir de un miARN.**

Esta es la búsqueda por defecto. El usuario puede realizar la búsqueda de genes blancos de miARNs conservados. En la primer pantalla se muestra los potenciales genes blancos para un miARN dado (Figura 4.8), con una breve descripción del gen, la familia a la que pertenece y además en cuantas y cuáles especies está presente. También, para cada especie detectada, se tiene acceso por pantalla al alineamiento del miARN-gen blanco, la energía de hibridación y los filtros empíricos de interacciones conocidas del par miARN-gen blanco (Figura 4.9).



Find potential miRNA targets

Find potential miRNA target families

Is this gene a potential miRNA target?

### Targets

miRNA: **miR398\_GTGTTCTCAGGTCACCCC**

MFE cutoff: **-29.232 kcal/mol**

MM Filter: **Yes**

Species: **All**

Go Back

Arabidopsis Tag <sup>[?]</sup>	Count <sup>[?]</sup>	Species <sup>[?]</sup>	Target description <sup>[?]</sup>	Gene family <sup>[?]</sup>	Alignments <sup>[?]</sup>
AT1G12520	16	Show/Hide	copper chaperone for SOD1		<a href="#">View</a>
AT1G08830	9	Show/Hide	copper/zinc superoxide dismutase 1		<a href="#">View</a>
AT3G15640	4	Show/Hide	Rubredoxin-like superfamily protein		<a href="#">View</a>
AT1G80230	3	Show/Hide	Rubredoxin-like superfamily protein		<a href="#">View</a>
AT2G26975	3	Show/Hide	Ctr copper transporter family		<a href="#">View</a>
AT2G40340	3	Show/Hide	Integrase-type DNA-binding superfamily protein	AP2-EREBP Transcription Factor Family	<a href="#">View</a>
AT3G27200	3	Show/Hide	Cupredoxin superfamily protein	Miscellaneous Membrane Protein Families	<a href="#">View</a>
AT3G43860	3	Show/Hide	glycosyl hydrolase 9A4	Glycoside Hydrolase Gene Families	<a href="#">View</a>
AT4G00050	3	Show/Hide	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	basic Helix-Loop-Helix (bHLH) Transcription Factor	<a href="#">View</a>
AT5G12220	3	Show/Hide	las1-like family protein		<a href="#">View</a>

**Figura 4.8: Resultados de la búsqueda de comTAR, con parámetros por defecto para el miR398**

Sequence ID	Species	5'-target-3' Alignment 3'-miRNA-5' [?]	MFE [?]
Aquca_013_00504.1	Aquilegia coerulea	TGGGCACCTGGGAACAT *   *       *       * CCCCACTGGACTCTTGTG	-31.7
471402	Arabidopsis lyrata	TGGGAGACCTGGGAACAC *   *           *        CCCCACTGGACTCTTGTG	-32.1
AT1G12520.1	Arabidopsis thaliana	TGGGAGACCTGGGAACAC *   *           *        CCCCACTGGACTCTTGTG	-32.1
Bradi5g18900.3	Brachypodium distachyon	TTGGTGACCTGGGAACGC **               *       * CCCCACTGGACTCTTGTG	-33.5
Bra026968	Brassica rapa	TGGGCACCTGGGAACAC *   *           *        CCCCACTGGACTCTTGTG	-32.5
Carubv10011816m	Capsella rubella	TGGGAGACCTGGGAACAC *   *           *        CCCCACTGGACTCTTGTG	-32.1
evm.model.supercontig_29.47	Carica papaya	TAGGTGACCTGAGAACAT **                   * CCCCACTGGACTCTTGTG	-34.2
Ciclev10021134m	Citrus clementina	TTGGTGACCTGGGAACAC **                        CCCCACTGGACTCTTGTG	-33.9
orange1.1g020436m	Citrus sinensis	TTGGTGACCTGGGAACAg **                       * CCCCACTGGACTCTTGTG	-32.7

Figura 4.9: **Identificación de genes regulados por un miARN en particular usando las herramienta comTAR** Captura de pantalla que muestra parte de la salida de comTAR del par miR398/SOD1 (At1g12520) en diferentes especies

**Realizar la búsqueda de familias de potenciales genes blancos de un miARN.**

Debido a que los miARNs en plantas en general regulan genes que codifican a proteínas de una misma familia, la herramienta tiene otra funcionalidad donde permite la búsqueda de genes agrupados por familias en vez de agruparlos por etiquetas. De este modo genes en distintas especies con diferentes etiquetas, pero que pertenecen a la misma familia pueden ser detectados como familias de potenciales genes blancos de miARNs conservados.

**Realizar la búsqueda de un gen de interés para ver si es potencial gen blanco de algún miARN conservado.**

El usuario puede introducir una etiqueta de locus en particular (tanto de Arabidopsis como el “gene ID” del Phytozome) y se identifica si este gen en particular puede ser un potencial gen blanco de algún miARN y además se muestra en cuantas especies se detecta. En Arabidopsis se utiliza el LocusID como identificador, mientras que en Phytozome este identificador varía según la especie y se puede ver la precedencia de cada especie en el sitio de Phytozome.

**Realizar la búsqueda de un miARN nuevo.**

En esta parte del programa el usuario puede realizar la búsqueda de nuevos ARNs pequeños teniendo en cuenta que la secuencia introducida tiene que ser de 18nt de largo (posiciones 2-19). Luego de la búsqueda, se da un link al usuario y después de unas horas, cuando haya sido procesado el cálculo, el usuario puede entrar a ese link y navegar los resultados por pantalla.

**4.3. Conclusiones.**

En este primer capítulo diseñamos una estrategia para identificar genes blancos regulados por miARNs en plantas, basado en la conservación evolutiva del par miARN-gen blanco. El enfoque requiere que la interacción miARN-gen blanco, pueda ocurrir en el contexto de un conjunto mínimo de parámetros que interactúan en diferentes especies pero la secuencia del gen blanco en sí, no necesariamente tiene que estar conservada. Además, nuestro enfoque permite ajustar el número de especies requeridas como un filtro para realizar la búsqueda con diferentes sensibilidades y relaciones señal/ruido.

Utilizando esta estrategia identificamos nuevos genes blancos potenciales en *A. thaliana*, a pesar de que este sistema ya había sido estudiado en detalles en distintos enfoques genómicos a gran escala ([2, 5, 53, 66, 112, 127]) y se validaron experimentalmente varios de ellos.

Varios de los nuevos genes blancos presentan funciones biológicas relacionadas con genes blancos ya conocidos. El miR159 regula factores de transcripción de la familia MYB y encontramos que también lo hace con NOZZLE (validado por nosotros en [35]). Ambos tipos de genes están involucrados en el desarrollo del estambre y polen[7, 98, 126, 161]. Vimos que el miR408 regula el transportador de cobre PAA2 (validado por nosotros en [35]), ya era conocido que lo hacía con proteínas de unión a cobre [53, 66, 127, 137]. Esto sugiere un rol general del miR408 en la homeostasis del cobre. El miR167 regula a la familia ARFs [118, 152] y vimos que también lo hacía con IAR3 (validado por nosotros en [35]). Estos genes participan en el control de niveles y actividad de auxinas [41, 150].

Además, tres de los nuevos genes blancos validados tienen “bulges” y los parámetros empíricos desarrollados por otros grupos de investigación usualmente le otorgan una gran penalidad a ellos, que puede llegar a ser el doble que un “mismatch” regular [66]. Sin embargo es probable que genes blancos con “bulges” asimétricos sean más frecuente de lo que se pensaba previamente en plantas. Es por esto que el enfoque ofrece una estrategia alternativa a otras predicciones que se basan en parámetros empíricos del par miARN-gen blanco [5, 38, 47, 66]. Una ventaja de la estrategia presentada es que las interacciones miARN-gen blanco conservadas probablemente participen en procesos biológicos relevantes. Además, esta estrategia puede ser fácilmente modificada para incorporar datos de otras bibliotecas, y/o para realizar la búsqueda de genes blancos presentes en un grupo específico de especies de plantas como se demostró para *Solanacea*.

Por último desarrollamos una herramienta web denominada comTAR para predecir potenciales genes blancos regulados por miARNs en plantas. Esta herramienta, es intuitiva y flexible, y se puede utilizar también para predecir potenciales genes blancos para nuevo ARN pequeños o para buscar familias de potenciales genes blancos de miARNs en plantas.

Esta parte del trabajo de Tesis fue publicado en la revista Nucleic Acid Research [35] y en la revista Bioinformatics [36]. La validación experimental de los genes blancos regulados por miARNs detectados con la estrategia, fueron realizadas por colegas del laboratorio. En particular, la validación de los genes blancos de *A. thaliana* por 5' RACE PCR modificada, las realizó Valeria Crosa. Y los experimentos en tabaco los realizó la Dra. Anabella Lodeyro.



## Capítulo 2.

### Estudios genómicos sobre la biogénesis de precursores de miARNs en plantas.

#### 5.1. Introducción.

En nuestro grupo se está estudiando la biogénesis de miARNs, específicamente como los precursores son reconocidos y procesados en plantas [23]. La biogénesis de los miARNs depende del reconocimiento de claves estructurales ubicadas en los precursores de miARN [15, 22, 38]. Estos precursores tienen estructuras de tallo-burbuja características [67], que se cree que proporciona las claves para el procesamiento del mismo y la liberación de los ARN pequeños de 21 nt.

Los precursores de miARNs en plantas son muy variables en tamaño y forma [23, 38]. Esa variabilidad se puede observar entre distintas familias de precursores, pero a veces también entre una misma familia de precursores en distintas especies (Figura 5.1).

Además, en las plantas, las mutaciones que impiden la biogénesis o actividad de los miARNs, tales como *hyll*, *hen1* y *ago1*, conducen al aumento en los niveles de los transcriptos de muchos de los genes blanco de miARNs (aproximadamente un 45 % de todos los blancos) [5, 26, 63, 160]. Esto sugiere que el mecanismo que involucra el corte y degradación de los ARNm es un componente importante de la represión inducida por los miARNs en plantas [67, 148].

En particular, la biogénesis de los miARNs es un proceso clave porque determina la secuencia exacta de nucleótidos del ARN pequeño funcional. En animales, los precursores poseen un tallo inferior de aproximadamente 11 pb, que se encuentra debajo de la posición donde DROSHA produce el corte. Luego del primer corte llevado a cabo por DROSHA, el pre-miARN liberado está conformado por el dúplex miARN/miARN\* y el loop terminal.

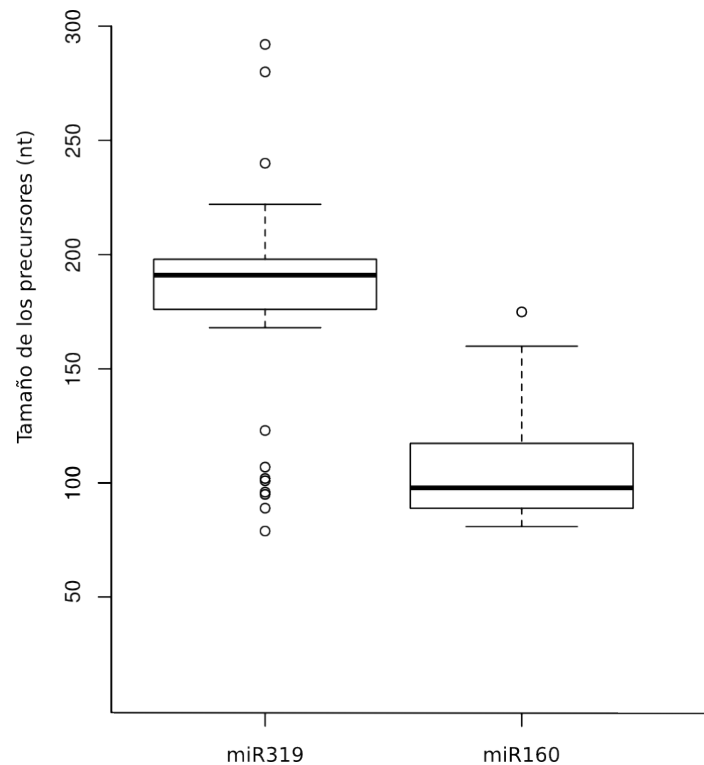


Figura 5.1: **Variación en el tamaño de precursores de miARNs de plantas.** Se muestra el tamaño (nt) de dos familias de precursores en distintas especies.

Estos pre-miARN son luego procesados a miARN maduros en el citoplasma mediante la interacción con la ribonucleasa Dicer.

Si bien en el caso de animales está claro cuáles elementos estructurales son reconocidos en los precursores durante su procesamiento, poco se sabía sobre el reconocimiento de los precursores de plantas por la maquinaria de procesamiento.

### 5.1.1. Construcción de bibliotecas de SPARE para estudios genómicos de biogénesis de miARNs en plantas.

En el marco de una colaboración con el grupo del Dr. Blake Meyers (Delaware, USA), el cual se especializa en secuenciación y análisis de ARN pequeños, nos propusimos entender cómo se procesan los precursores de miARNs en plantas. Colegas del laboratorio realizaron una estrategia para analizar sistemáticamente intermediarios de procesamiento de miARNs y caracterizar la biogénesis de los miARNs presentes en *A. thaliana* mediante técnicas de secuenciación de alto rendimiento, utilizando los equipos de última generación disponibles en Delaware (USA). Esta técnica desarrollada en el laboratorio se conoce como SPARE

[124] (del inglés Specific Parallel Amplification of RNA Ends). La estrategia utilizada a partir de los datos de bibliotecas de SPARE se muestra en la Figura 5.2.



Figura 5.2: Esquema del procedimiento para analizar los datos de SPARE.

La técnica de SPARE, permite identificar los intermediarios de procesamiento de precursores de miARNs, y uniendo la información brindada por los diferentes fragmentos es posible dilucidar tanto la dirección, como el número de cortes requeridos para la biogénesis de cada miARN (Figura 5.3). Es decir que no sólo se puede detectar la dirección de procesamiento, sino que también la técnica permite identificar precursores que requieren más de dos cortes para liberar el miARN maduro. Se utilizó esta técnica de SPARE para determinar el mecanismo de procesamiento cada uno de los precursores de *A. thaliana*.

### 5.1.2. Visualización de precursores que se procesan desde la base.

En el caso de los precursores que se procesan de base a loop, el único corte que puede ser detectado es el corte proximal (Figura 5.4). En la Figura se puede ver el primer corte producido por DCL1 de los precursores del miR168a, miR172a y miR395b. Las flechas indican la posición en la que terminan los fragmentos secuenciados y, debajo de la misma, se representa el número de veces que ese fragmento se secuenció, y por lo tanto indica la abundancia de ese fragmento en la célula. Se puede observar que este corte se da en una región de ARNdh ubicada a  $\sim 15$  nt desde el loop interno. Por lo tanto se concluye que estos precursores de miARN tienen un tallo de ARNdh inferior a la región donde se ubica el propio

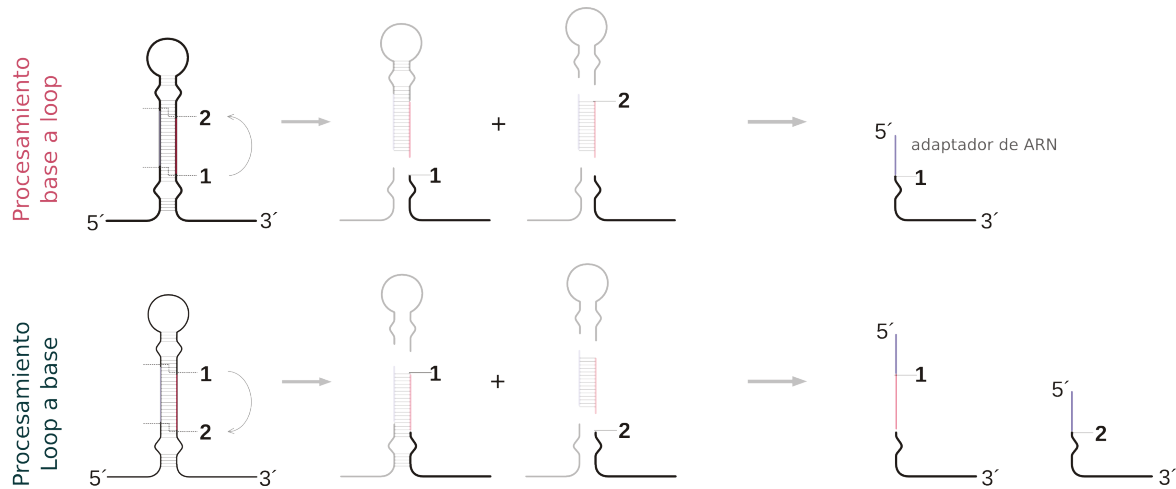


Figura 5.3: **Técnica de SPARE para diferenciar mecanismos de procesamiento de precursores.** Detalle donde se muestra a nivel molecular como la técnica permite diferenciar entre dos direcciones de procesamiento opuestas. Mediante la técnica se puede detectar un solo corte en los precursores procesados de base a loop. Mientras que para los precursores procesados de loop a base, la técnica permite detectar todos los cortes.

miARN. Puede verse que la gran mayoría de los cortes están ubicados en este sitio (Figura 5.4).

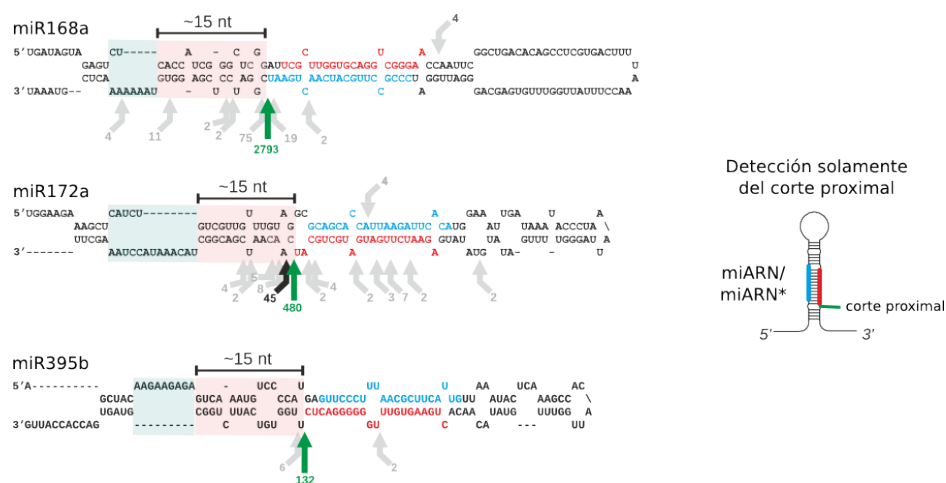


Figura 5.4: **Identificación y caracterización de precursores de miARNs procesados de base a loop.** Esquema donde se muestra la estructura secundaria del miR168a, miR172b y miR395b. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestra el corte más abundante, que también coincide con al corte proximal del miARN/miARN\*. Flechas en negro muestran otros cortes con al menos 5% de abundancia del número total de cortes, mientras que otros cortes minoritarios se muestran con una flecha gris. En rosa se resalta el tallo de 15nt debajo del corte proximal. En celeste se resalta el "loop" interno. El miARN se indica en color rojo y el miARN\* en color azul. La imagen de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

### 5.1.3. Visualización de precursores que se procesan desde el loop.

Para los precursores que se procesan de loop a base, la técnica de SPARE permite detectar tanto el corte distal como el proximal, que son los necesarios para liberar al miARN maduro (Figura 5.5). Además, en los precursores que necesitan más de dos cortes para liberar el miARN maduro, como la familia del miR319, los cuatro cortes pueden ser detectados por esta técnica. En la figura se pueden ver los dos cortes por DCL1 de miembros de la familia del miR156 y del precursor del miR160a, que son procesados desde el loop a la base.

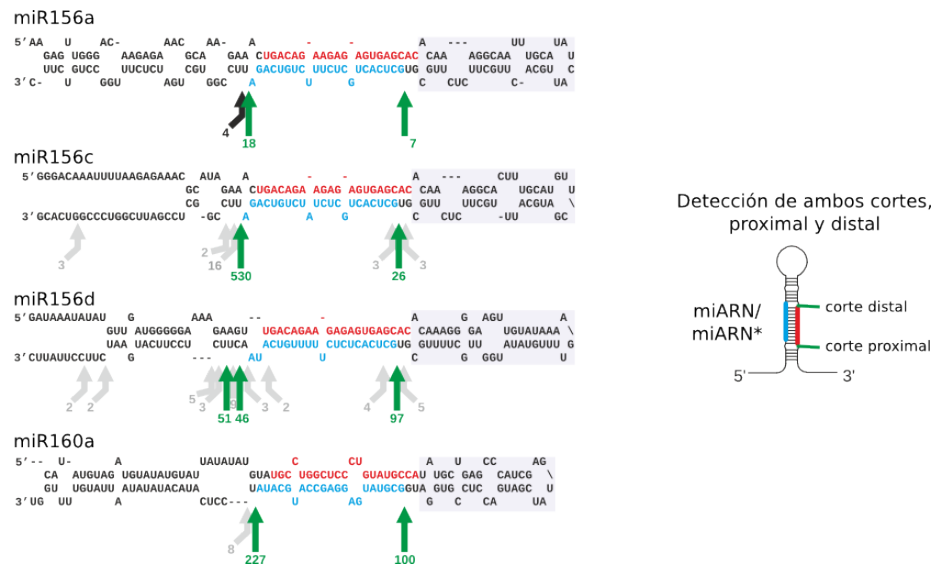


Figura 5.5: **Identificación y caracterización de precursores de miARNs procesados de loop a base.** Esquema donde se muestra la estructura secundaria del miR156a, miR156c, miR156d y miR160a. Las flechas indican la posición y número de lecturas de los cortes del precursor identificado. Flechas en verde muestra el corte más abundante, que también coincide con al corte proximal del miARN/miARN\*. Flechas en negro muestran otros cortes con al menos 5% de abundancia del número total de cortes, mientras que otros cortes minoritarios se muestran con una flecha gris. Con gris se resalta el tallo superior del miR156 y miR160. El miARN se indica en color rojo y el miARN\* en color azul. La imagen de la derecha muestra el patrón de corte típico detectado en la biblioteca de SPARE para estos precursores.

## 5.2. Resultados y Discusión.

### 5.2.1. Desarrollo de herramienta para el análisis de intermediarios de procesamiento en *A. thaliana*.

#### Análisis de las bibliotecas de SPARE y diseño de la herramienta.

En una primera serie de experimentos, Nicolás Bologna y Arnaldo Schapire realizaron bibliotecas de SPARE en Col-0 y mutantes en *fiery* para estudiar los intermediarios de procesamiento en *A. thaliana*. De esta técnica, se obtienen una gran cantidad de datos para cada biblioteca secuenciada y los primeros análisis se hicieron manualmente.

Por lo tanto, se decidió en el laboratorio generar un protocolo bioinformático para el análisis de bibliotecas SPARE. Como material de partida, Belén Moro realizó nuevas bibliotecas de SPARE para precursores de *A. thaliana* de plantas silvestres (Col-0) y en *hyll*, *se* y *fiery*. El procesamiento y análisis de los datos se hizo a partir de los datos crudos y se detallan en Materiales y Métodos (ver sección 3.3). El análisis de estas nuevas bibliotecas permitió estudiar a plantas mutantes y también permitió aumentar el número de precursores de miARNs de Arabidopsis caracterizados. La tabla completa con la dirección de procesamiento de miARNs conservados en *A. thaliana* determinados a partir de los resultados de las bibliotecas de SPARE realizadas por Schapire, Bologna y Moro, se muestra en el Anexo (Tabla ).

Para ayudar a analizar estos datos complejos, desarrollamos una herramienta intuitiva para el análisis y visualización de los datos obtenidos por la técnica de SPARE. Esto nos permite facilitar el análisis de la identificación de los intermediarios de procesamiento en de este tipo de bibliotecas. Brevemente, en total se obtuvieron cerca de 150 millones de secuencias sumando todas las bibliotecas (Tabla 5.1). Los datos de la secuenciación son procesados y las secuencias correspondientes a fragmentos de precursores son agrupados y mapeados contra la totalidad de los precursores de miARN de *A. thaliana*.

Esa información ya procesada es la que utilizamos en la herramienta para analizar y comparar los intermediarios de procesamiento para plantas silvestres y para las mutantes de procesamiento.

La herramienta diseñada permite seleccionar un precursor en particular y luego seleccionar una o varias mutantes y controles que se quieren analizar. Además, la herramienta permite

Tabla 5.1: Bibliotecas de SPARE

Bibliotecas	Muestras	Secuencias totales	Secuencias que mapean los precursores	Secuencias únicas que mapean los precursores
col_2AB	Col-0 réplica 1. Control de fiery y hyl1	13911694	80166	308
col_3AB	Col-0 réplica 2. Control de fiery y hyl1	16618008	126556	426
Col_AD	Col-0 réplica 1. Control de se	13758567	119368	496
Col_BC	Col-0 réplica 2. Control de se	14648459	241973	553
FIERY_1AB	fiery réplica 1	9832923	470789	1655
FIERY_4	fiery réplica 2	23529725	821562	1752
HYL_1	hyl1 réplica 1	10171629	45653	316
HYL_2	hyl1 réplica 2	8864406	35860	320
SE_BD	se réplica 1	15291993	299513	639
SE_CE	se réplica 1	25296809	510438	693

modificar otros parámetros como la longitud y abundancia de los fragmentos secuenciados a considerar.

### Precursores procesados desde la base.

Primero, vemos la salida de la herramienta para el precursor del miR165a, analizando distintas mutantes de procesamiento al mismo tiempo (Figura 5.6). La posición final del dúplex miARN/miARN\* fue considerada como la posición 0, y las posiciones positivas son bases hacia el loop del dúplex mientras que las posiciones negativas son bases hacia abajo del dúplex. Los fragmentos se muestran en forma porcentual tomando cada biblioteca de forma independiente y en distintos colores se representan las distintas bibliotecas. Además, se muestra una tabla con los valores absolutos de cada fragmento y con flechas verdes se marcan las posiciones con los fragmentos de mayor abundancia en promedio de todas las bibliotecas (Figura 5.6).

En la Figura 5.6 se muestra un precursor que es procesado con un mecanismo corto de base a loop, donde según se muestra en la Figura 5.3, los fragmentos detectados corresponden únicamente al primer corte por DCL1. Se pueden observar también cortes en las posiciones  $\pm 2$  con respecto a las esperadas por actividad “sloopy” (poco rigurosa) de DCL1 [123]. Además, se observa que en la posición -5 aparecen fragmentos en todas las bibliotecas aunque en muy baja abundancia (Figura 5.6). Por el momento, no hay una clara explicación para el origen mecanístico de estos cortes, pero muestra la utilidad de la herramienta para poder fácilmente visualizar e identificar elementos comunes e inesperados en distintas bibliotecas.

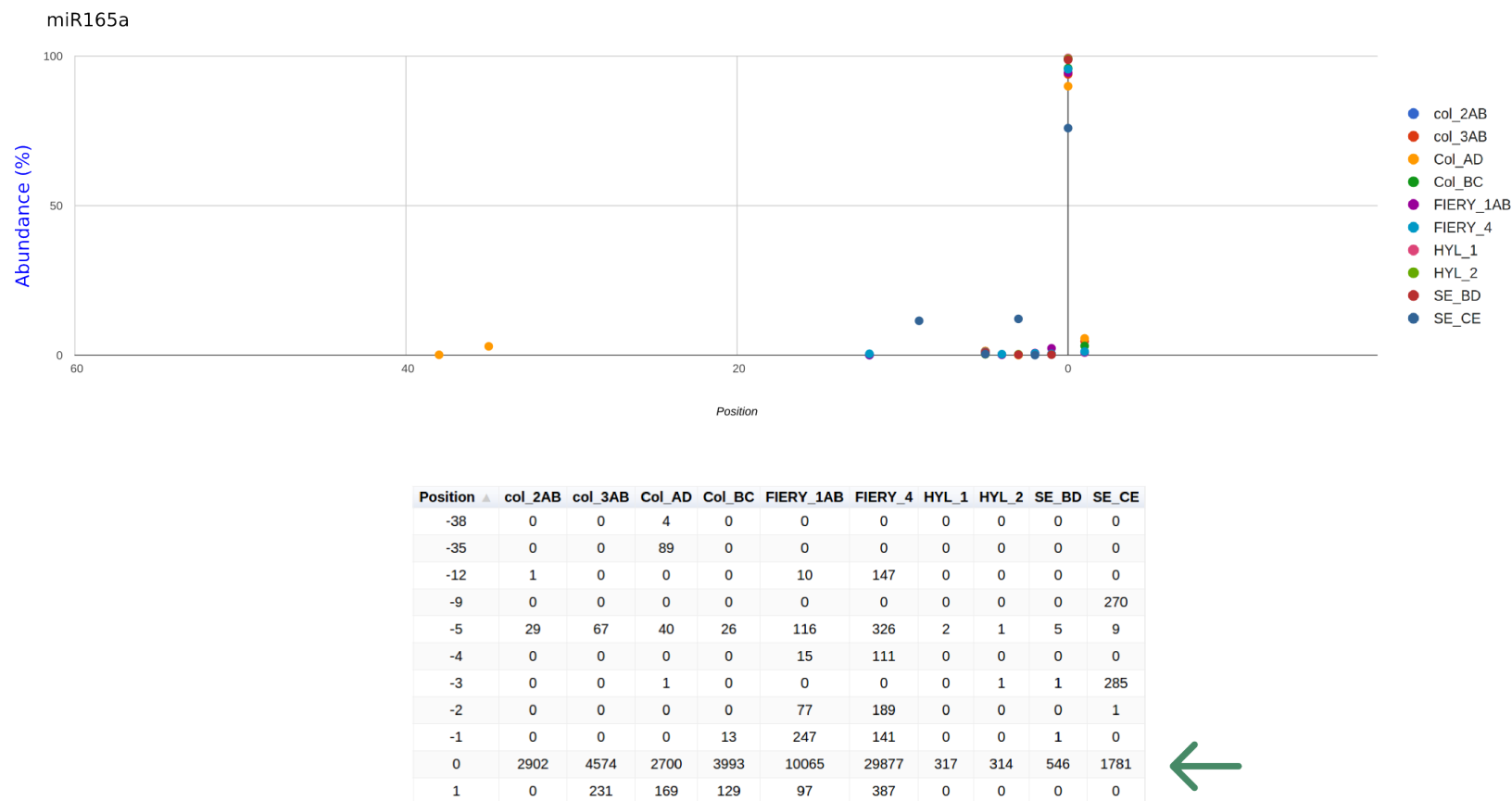
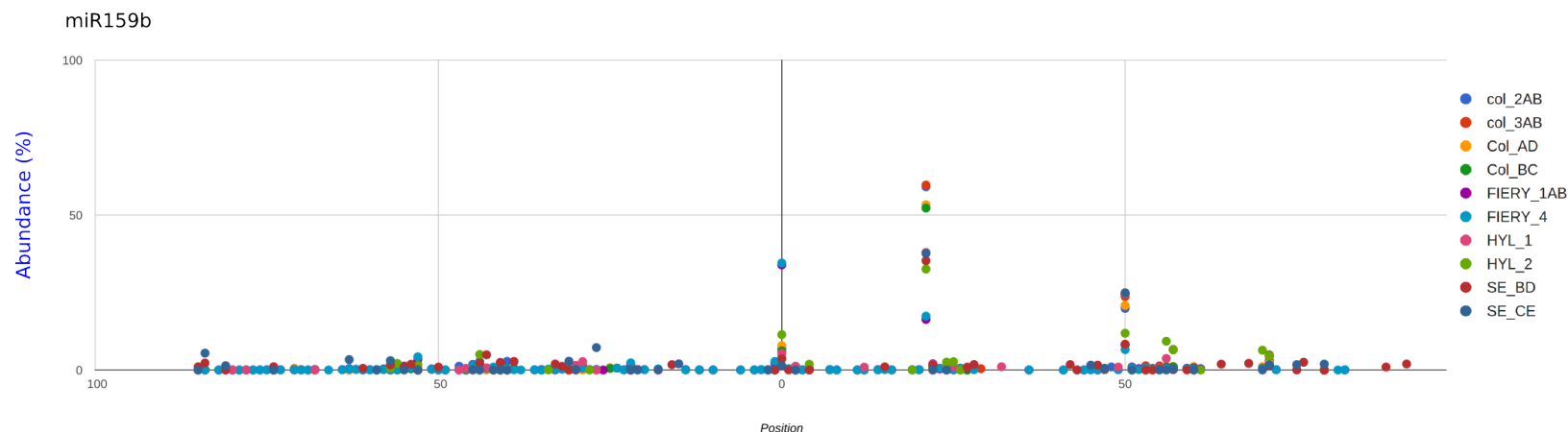


Figura 5.6: **Captura de pantalla de la herramienta de SPARE para el miR165a.** Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos en esa posición dividido la abundancia total de los fragmentos en el precursor). La posición final del dúplex miARN/miARN\* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.



**Precursores procesados desde el loop.**

Para el caso de los precursores cortos de loop a base, los fragmentos del corte por DCL1 detectados, pueden ser más de uno según muestra la técnica de SPARE (Figura 5.3). Lo mismo sucede con los precursores de loop a base que son cortados más de dos veces por DCL1. Para ilustrar esto, mostramos el caso del miR159, que al igual que todos los precursores que son procesados secuenciales de loop a base, los cuatro fragmentos más abundantes corresponden a los cuatro cortes realizados por DCL1 que son requeridos para procesar este precursor (Figura 5.7). El primer corte se da en la posición 71, que de los cuatro cortes es el que tiene fragmentos menos abundantes, y luego DCL1 corta a 21 nucleótidos liberando otros ARN pequeños. Luego se pueden observar los fragmentos correspondientes al tercer corte (posición 21) y cuarto corte (posición 0), que son los necesarios para liberar el miARN maduro (Figura 5.7). En este caso se pueden ver otros fragmentos, aunque nunca son tan abundantes como los correspondientes a los cortes por DCL1. Estos otros cortes posiblemente sean el resultado del corte improductivo del precursor por ribonucleasas y quizás sean parte de un proceso de degradación de los precursores. Los orígenes de estos fragmentos del precursor están en estudio en el laboratorio. Por cuestiones de simplicidad, en la tabla solo se muestra los fragmentos con mayor abundancia en promedio de todas las bibliotecas.



Position	col_2AB	col_3AB	Col_AD	Col_BC	FIERY_1AB	FIERY_4	HYL_1	HYL_2	SE_BD	SE_CE
21	10252	19917	5562	10992	6776	20171	2514	1477	1685	2752
50	3450	7908	2174	5164	3392	7627	1629	538	398	1824
0	1209	1319	822	1300	14091	40065	353	518	172	101
71	451	1070	233	1028	538	1713	302	201	70	125
53	9	461	1	64	250	822	0	0	1	0
55	92	450	16	157	365	639	52	24	17	1
57	32	379	66	218	115	227	442	294	6	23



Figura 5.7: **Captura de pantalla de la herramienta de SPARE para el miR159b.** Porcentaje de fragmentos del precursor (abundancia relativa de los fragmentos en esa posición dividido la abundancia total de los fragmentos en el precursor). La posición final del dúplex miARN/miARN\* fue considerada como la posición 0. La tabla muestra los valores absolutos de cada fragmento. Por cuestiones de simplicidad, solo se muestran en la tabla los fragmentos con mayor abundancia en promedio de todas las bibliotecas. Las flechas verdes marcan las posiciones del precursor con los fragmentos de mayor abundancia.

### 5.2.2. Análisis estructural de precursores de miARNs en plantas.

Una vez que se pudo clasificar a los precursores de acuerdo a su tipo de procesamiento, ya sea desde el loop o desde la base, y sea con dos o mas cortes de DCL1, decidimos ver si había diferencias estructurales para cada grupo. Los datos para este primer análisis fueron el SPARE realizado sobre las plantas silvestres y mutantes *fiery*.

Primeramente, obtuvimos las estructuras secundarias para cada precursor utilizando el programa Mfold[169] (Ver Materiales y Métodos sección 3.5) y realizamos un análisis estructural conjunto de los precursores correspondientes a los dos mecanismos de procesamiento, los que se procesan en dirección base a loop (Figura 5.8) y los que se procesan loop a base (Figura 5.9). Definimos a una coincidencia en cada posición con un 0, mientras que “bulges” y “mismatches” los consideramos como 1. El lado proximal del dúplex miARN/miARN\* fue definido como la posición +1 y analizamos desde la posición -25 a la posición +40 (Figura 5.8 y 5.9).

#### Estructuras secundarias de precursores procesados desde la base.

El análisis de los precursores que se procesan de base a loop mostró que todos ellos tienen un claro tallo inferior de 15 nt de largo (Figura 5.8). Este tallo se pudo observar tanto para los precursores validados experimentalmente en esta parte del trabajo, como para todos los precursores conservados que se procesan desde la base (Figura 5.8 en violeta). También, pudimos observar que las bases inmediatamente debajo del dúplex miARN/miARN\* (posición -2 y -1) tienden a estar desapareadas (Figura 5.8). Además, las posiciones -3 y -4 y las 3 últimas posiciones del tallo inferior (-13,-14 y -15) están apareadas casi siempre (Figura 5.8). En general, nuestros resultados muestran que los precursores procesados en una dirección base a loop son más uniformes de lo que se pensaba previamente.

#### Estructuras secundarias de precursores procesados desde el loop.

Con la excepción de dos miARNs (miR396a y miR162b) estos precursores procesados desde el loop, no tienen una estructura obvia debajo del dúplex miARN/miARN\* (Figura 5.9). En cambio, estos precursores tienen una región terminal estructurada (Figura 5.9), que tiene un tamaño homogéneo de 42nt que incluye un loop corto en contraste con la misma región en los precursores que se procesan de base a loop donde es más variable (Figura 5.8 y 5.9).

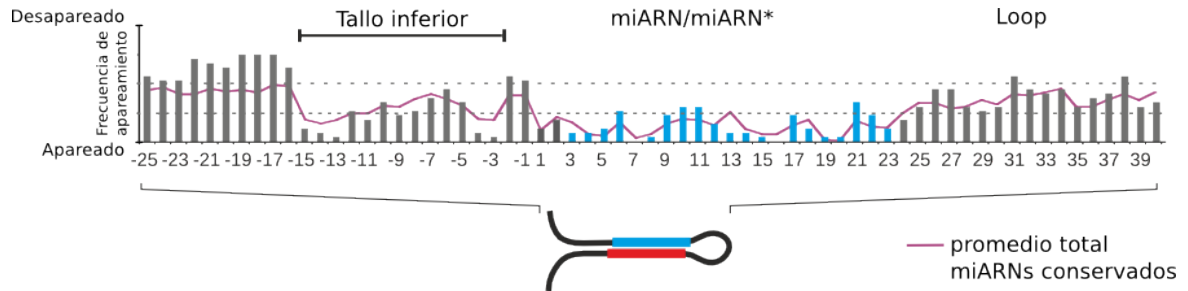


Figura 5.8: **Estructura secundaria de precursores detectados que se procesan en dirección base a loop.** Los matches en cada posición los consideramos como 0, mientras que “bulges” y “mismatches” fueron considerados como 1. La estructura secundaria considerando todos los miARNs conservados se indica con color violeta.

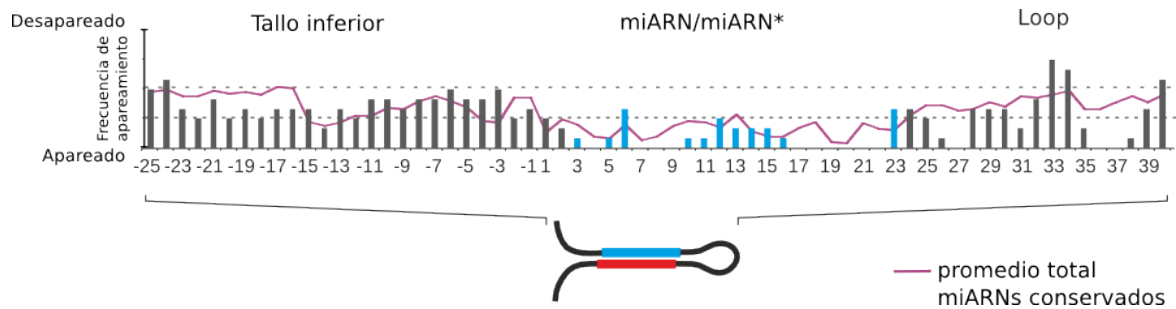
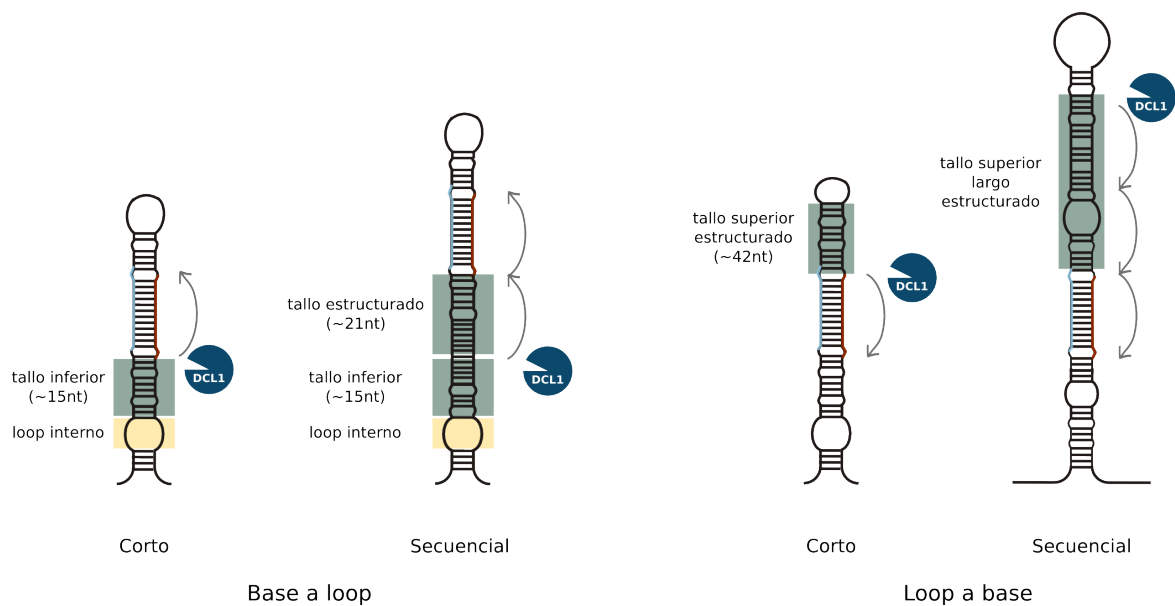


Figura 5.9: **Estructura secundaria de precursores detectados que se procesan en dirección loop a base.** Los matches en cada posición los consideramos como 0, mientras que “bulges” y “mismatches” fueron considerados como 1.

### 5.3. Conclusiones.

En esta segunda parte del proyecto de Tesis presentamos una estrategia y para el análisis sistemático de las bibliotecas de SPARE y la identificación de la biogénesis de precursores de miARNs desde un punto de vista genómico. Utilizando esta técnica encontramos que los miARNs son procesados por cuatro mecanismos, dependientes de la dirección secuencial de la maquinaria de procesamiento y del número de cortes requeridos para liberar el miARN (Figura 5.10). Y con esto, pudimos determinar la dirección de procesamiento para la mayoría de los miARNs conservados en *A. thaliana* (Anexo, Tabla ).



**Figura 5.10: Distintos mecanismos de procesamiento de miARNs en plantas** A la izquierda se muestra los precursores que son procesado de base a loop. A la derecha se muestra los precursores que son procesado de loop a base. En ambos casos se muestran los precursores cortos y los secuenciales. Se identifica en azul el corte por DCL y con flechas la dirección de procesamiento.

La clasificación de los precursores, teniendo en cuenta los mecanismos de procesamiento y el análisis consiguiente para cada grupo, reveló determinantes estructurales específicos para cada grupo.

Integrando los resultados que se muestran en la introducción junto a los resultados obtenidos en esta Tesis, se pudo demostrar que los precursores de miARNs en plantas se pueden agrupar por cuatro mecanismos de procesamiento con distintas características (Figura 5.10).

- En los precursores con un mecanismo **corto de base a loop**, un loop interno seguido por un tallo inferior de  $\sim 15$ nt especifica la posición del primer corte. Esta estructura se encuentra en la mayoría de familias de miARNs [94, 129, 151]. A pesar de que el tallo puede contener bulges, la transición de un loop interno (simple hebra) al tallo inferior es bastante marcada, y tres pares de bases apareadas generalmente definen el comienzo del tallo inferior del precursor [23]. El segundo corte procede a una distancia fija de  $\sim 21$  nt desde la posición del primer corte.
- En los precursores con un mecanismo **secuencial de base a loop** (ej: familia del miR169), el primer corte procede como en los cortos de base a loop, pero luego son necesario dos cortes más para liberar el miARN, generando en el proceso niveles bajos de RNA pequeños adicionales [23].
- En los precursores con un mecanismo **cortos de loop a base** (ej: familia del miR156 y miR160), el procesamiento es guiado por un tallo superior, y son necesarios dos cortes para liberar el miARN maduro. La región terminal de estos precursores tienen una largo conservado de  $\sim 42$  donde incluye un loop pequeño [23].
- En los precursores con un mecanismo **secuencial de loop a base** (ej: familia del miR319 y miR159), cuatro cortes secuenciales por DCL1 son los encargados de procesar los precursores de miARNs. En general muestran un tallo largo superior, del cual otros ARNs pequeños son generados [4, 21, 23]

Estos resultados fueron publicado en la revista Genome Research [23]. Los resultados ofrecen una explicación para la diversidad estructural de los genes de precursores de miARN en plantas y nuevas perspectivas hacia la comprensión de la biogénesis de los ARNs pequeños [23].

Además, pudimos encontrar determinantes estructurales distintos para precursores que se procesados desde la base y precursores procesados desde el loop.

## Capítulo 3.

### Enfoque bioinformático para el estudio de la evolución y biogénesis de miARNs en plantas.

#### 6.1. Introducción.

En el capítulo anterior vimos que los precursores de plantas pueden ser procesados por cuatro mecanismos con distintas características (Figura 5.10). Esto contribuye a la complejidad del sistema de procesamiento, respecto al de animales.

Los estudios que dan origen a estos modelos de procesamiento se basan principalmente en datos de secuenciación masiva de ARN en los que se detectan los intermediarios de procesamiento de los distintos precursores. Como se mostró anteriormente, a partir de los intermediarios de procesamiento es posible definir cuál es el mecanismo al que es sometido cada precursor.

En estudios previos que consideran a los precursores de miARNs en plantas como un solo grupo de moléculas se han encontrado pocos elementos comunes en sus estructuras [94]. En este capítulo, aprovechamos los datos genómicos generados por nuestro laboratorio a partir del estudio del procesamiento de miARNs en plantas, y agrupamos los precursores por mecanismo de procesamiento y estudiamos en detalle a cada precursor en forma individual. Además, incluimos en el análisis distintas especies de plantas para poder estudiar el procesamiento de miARNs mediante un enfoque evolutivo.

Que los precursores de miARNs en plantas sean procesados de diferentes maneras [23], nos llevó a especular que su patrón de evolución también puede ser diferente y podrían estar vinculados a su mecanismo de procesamiento. El objetivo es deducir los patrones de los precursores de plantas que sirvan de guía durante el proceso de biogénesis de miARNs. Es por esto, que elaboramos una estrategia bioinformática para caracterizar la relación entre

la evolución de los precursores de miARNs en plantas y los mecanismos de procesamiento determinados previamente.

## **6.2. Resultados y Discusión.**

### **6.2.1. Estrategia bioinformática para el estudio de la evolución y biogénesis de miARNs en plantas.**

La metodología utilizada para el estudio de la evolución y biogénesis de miARNs en plantas se detalla en la Materiales y Métodos (sección 3.6). En una primera parte, la utilizamos para estudiar precursores de miARNs conservados en dicotiledóneas. Brevemente, la misma consta de los siguientes pasos:

- Búsqueda de ortólogos para cada miembro de cada familia de Arabidopsis de precursores de miARNs conservados en plantas.
- Extensión de las secuencias.
- Alineamiento de secuencia primaria.
- Alineamiento de secuencia teniendo en cuenta estructura secundaria.
- Búsqueda de motivos conservados.
- Representación gráfica de los precursores.

#### **Búsqueda de ortólogos de precursores de miARNs.**

Las secuencias de los precursores de miARNs en miRBase<sup>14</sup> (base de datos biológica que contiene secuencias y anotaciones de miARNs) rara vez son validadas experimentalmente, y de hecho muchas de ellas son el resultado de predicciones computacionales de estructuras en forma de tallo y burbuja que incluye el miARN maduro, pero que no se extiende muchas bases más allá del mismo [72]. Por esto, los determinantes estructurales pueden no estar incluidos en las secuencias mostradas como precursores en miRBASE, en especial si son procesados desde la base. Por lo tanto, empezamos nuestro análisis con una definición arbitraria de los precursores de plantas incluyendo 150 nt fuera del par miARN/miARN\*. En

---

<sup>14</sup><http://mirbase.org>



ese sentido, extendimos la secuencia de los precursores para el posterior análisis utilizando datos genómicos de plantas extraídos de Phytozome<sup>15</sup> para 30 especies de dicotiledóneas.

En segundo lugar, para cada miembro de cada familia de *A. thaliana* no es trivial asignarle un ortólogo en otra especie teniendo en cuenta la anotación de miRBase. Por esto, realizamos una búsqueda de ortólogos para cada miembro de cada familia de *A. thaliana* utilizando como criterio la técnica de Blast recíproco. Esta técnica consiste en utilizar Blast cruzado para predecir ortólogos putativos (Ver Materiales y Métodos sección 3.6).

Algunas familias de miARNs se han diversificado durante la evolución, donde en algunas especies un miARN puede encontrarse como un único gen y en otras especie el mismo miARN puede encontrarse en múltiples miembros. Por ejemplo, el miR396 en Arabidopsis está constituido por dos miembros, miR396a y miR396b mientras que en el caso de arroz existen 8 miembros, nombrados del miR396a al miR396h. Esta variación en el número de genes de miARNs en distintas especies implica que no necesariamente existe un ortólogo a un determinado miARN de Arabidopsis en cada una de las especies. Sin embargo, no podemos descartar que estemos perdiendo de encontrar miARNs ortólogos por la propia limitación en la búsqueda que hacemos utilizando el blast recíproco.

### **Alineamiento de los precursores en base a su secuencia primaria.**

Como resultado de la búsqueda de ortólogos de precursores de miARNs de los 93 precursores de Arabidopsis encontramos 1819 precursores en total en las 30 especies de dicotiledóneas. En la Figura 6.1 se muestran las estructuras secundarias de los ortólogos del precursor de miR172a utilizando el programa RNAfold [87]. Como puede verse, si bien hay una similitud entre las estructuras es difícil deducir información concreta a partir de esta Figura.

A continuación, comenzamos con los alineamientos múltiples de cada miembro de las familias de precursores de miARNs conservadas en dicotiledóneas. Primero, realizamos el alineamiento en base a su secuencia primaria utilizando la herramienta T-Coffee [103].

En la Figura 6.2 se muestra el alineamiento del precursor del miR172a en distintas especies, coloreado en base a la conservación evolutiva de la secuencia primaria. Se puede observar que el miR172a maduro y el miR172a\* están muy conservados en las distintas especies, pero además se puede ver una cola de conservación hacia la izquierda del miARN y hacia la derecha del miARN\* (Figura 6.2 en líneas de puntos).

---

<sup>15</sup><http://phytozome.jgi.doe.gov>

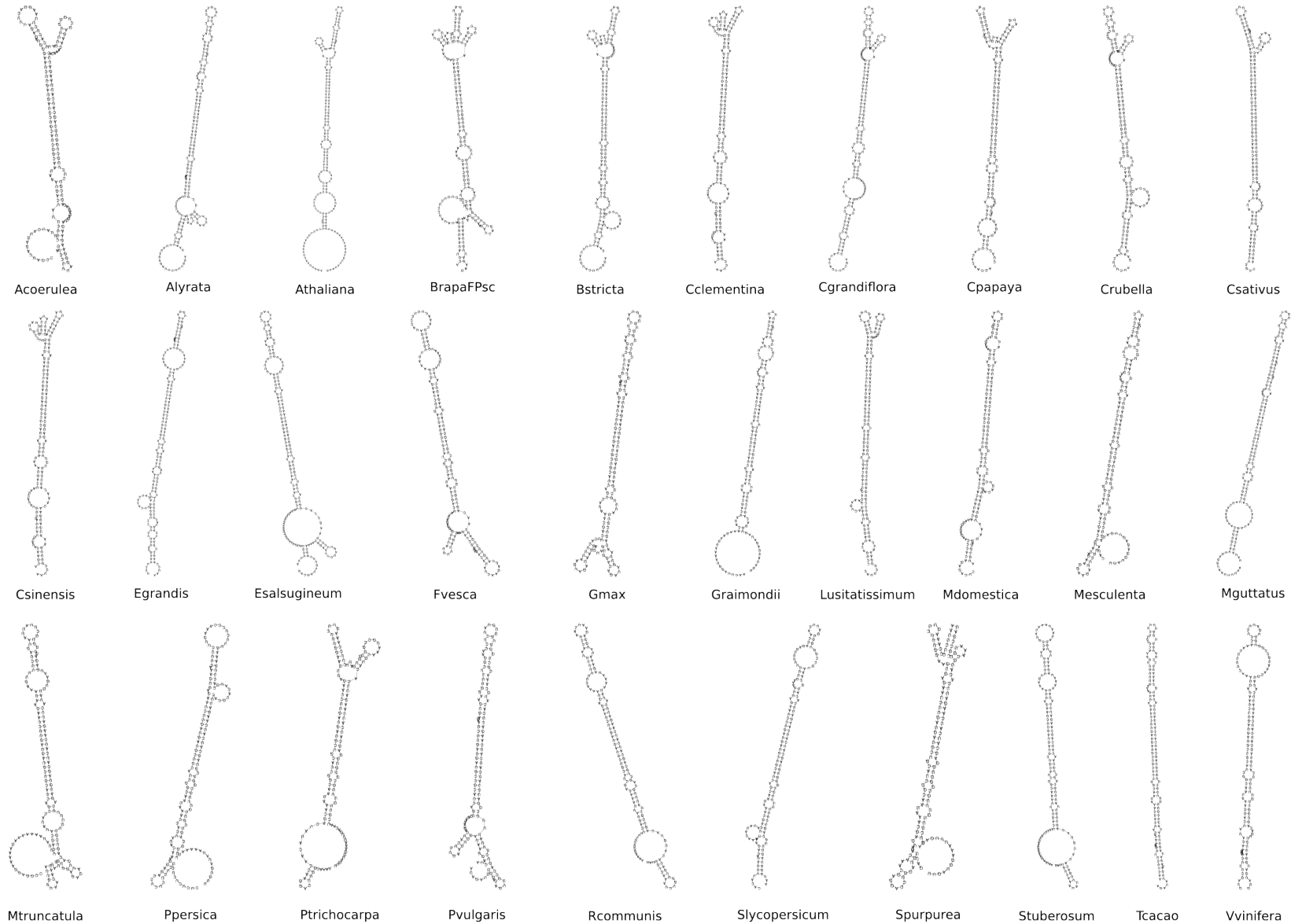


Figura 6.1: **Estructura secundaria del miR172a en distintas especies.** Se muestra la estructura secundaria de los precursores del miR172a en distintas especies, calculados con la herramienta RNAfold. Se puede observar que existe un patrón estructural que comparten los precursores, en la región inmediata por debajo del dúplex miARN/miARN\*.

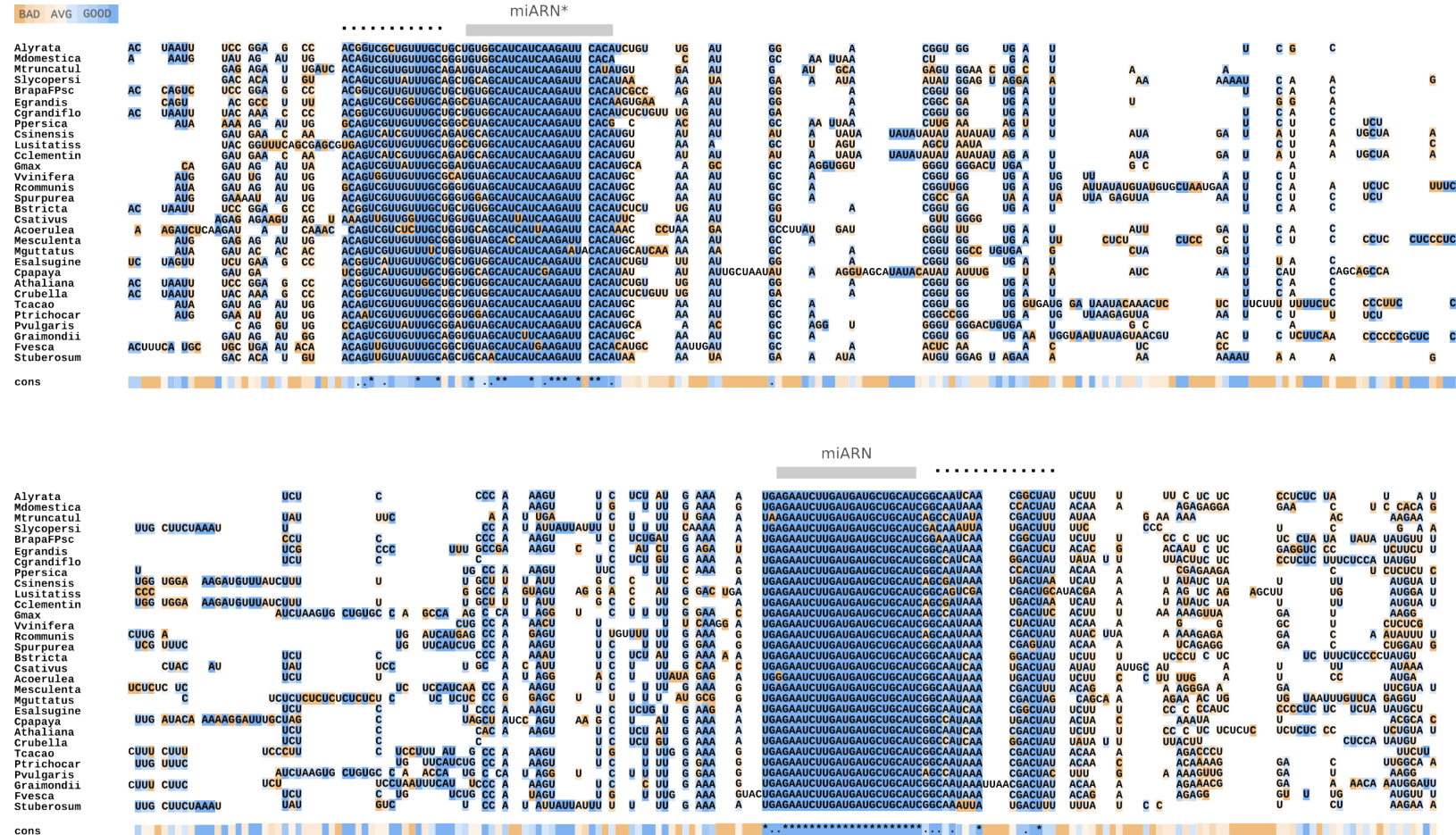


Figura 6.2: **Alineamiento del precursor del miR172a.** Se muestra el alineamiento del precursor del miR172a en distintas especies. En colores se muestra la conservación de la secuencia primaria, donde azul más oscuro denota mayor conservación y naranja menor conservación. Se muestran ~60 bases por debajo del dúplex miARN/miARN\* y los alineamientos están separados en dos líneas para una mejor representación. Se puede observar conservación en el miARN y el miARN\*. Además, se marca en línea de puntos una cola de conservación.

Esta cola de conservación se corresponde al tallo inferior del precursor. Sin embargo, esta forma de visualización no hace fácil correlacionar conservación con estructura secundaria.

Después de estos primeros análisis, llegamos a la conclusión de que el análisis de la conservación evolutiva de cada precursor era valioso ya que permitía identificar regiones conservadas que tienen una función en el procesamiento. Sin embargo, también notamos que la forma de representación de esta información no era del todo conveniente, en especial, teniendo en cuenta la cantidad de precursores estudiados en esta parte del trabajo de Tesis.

### **Diseño de una implementación gráfica para la visualización integral de los precursores.**

Para poder reconocer fácilmente la región conservada dentro del precursor y poder analizar más en detalle estos patrones de conservación, decidimos agrupar toda esta información generada a partir de los alineamientos de secuencia primaria y estructura secundaria en un gráfico circular como se muestra en la Figura 6.3. Estos gráficos se hicieron utilizando el paquete Circos [74], que adaptamos para visualizar los precursores de miARNs.

En dicha Figura, se muestra el Circos del precursor del miR172a a modo de ejemplo. Para representar el precursor, se tomaron 60 nt por debajo del dúplex miARN/miARN\*. En el anillo exterior se representa, en colores azules y marrones, el grado de conservación de la secuencia consenso a partir de los alineamientos en base a su secuencia primaria. Además, se muestran las bases del precursor de *A. thaliana* que corresponden a la posición consenso. Luego en el anillo interior se representa, mediante un histograma, la frecuencia de bases apareadas y desapareadas para cada posición del precursor en las distintas especies, donde una barra verde indica que la posición tiende a estar apareada formando una estructura secundaria, mientras que la barra violeta indica que la base tiende a estar desapareada. Con líneas de distinto espesor se muestra la interacción entre pares de bases del precursor considerando la estructura secundaria. Finalmente por fuera del anillo exterior se marca la secuencia del miARN y miARN\*.

Estos gráficos denominados Circos, los adaptamos y los realizamos, para visualizar de manera simple los precursores de miARNs en distintas especies de plantas. Y luego fueron utilizados para caracterizar la relación entre los patrones de conservación y los mecanismos de procesamiento en plantas determinados previamente.

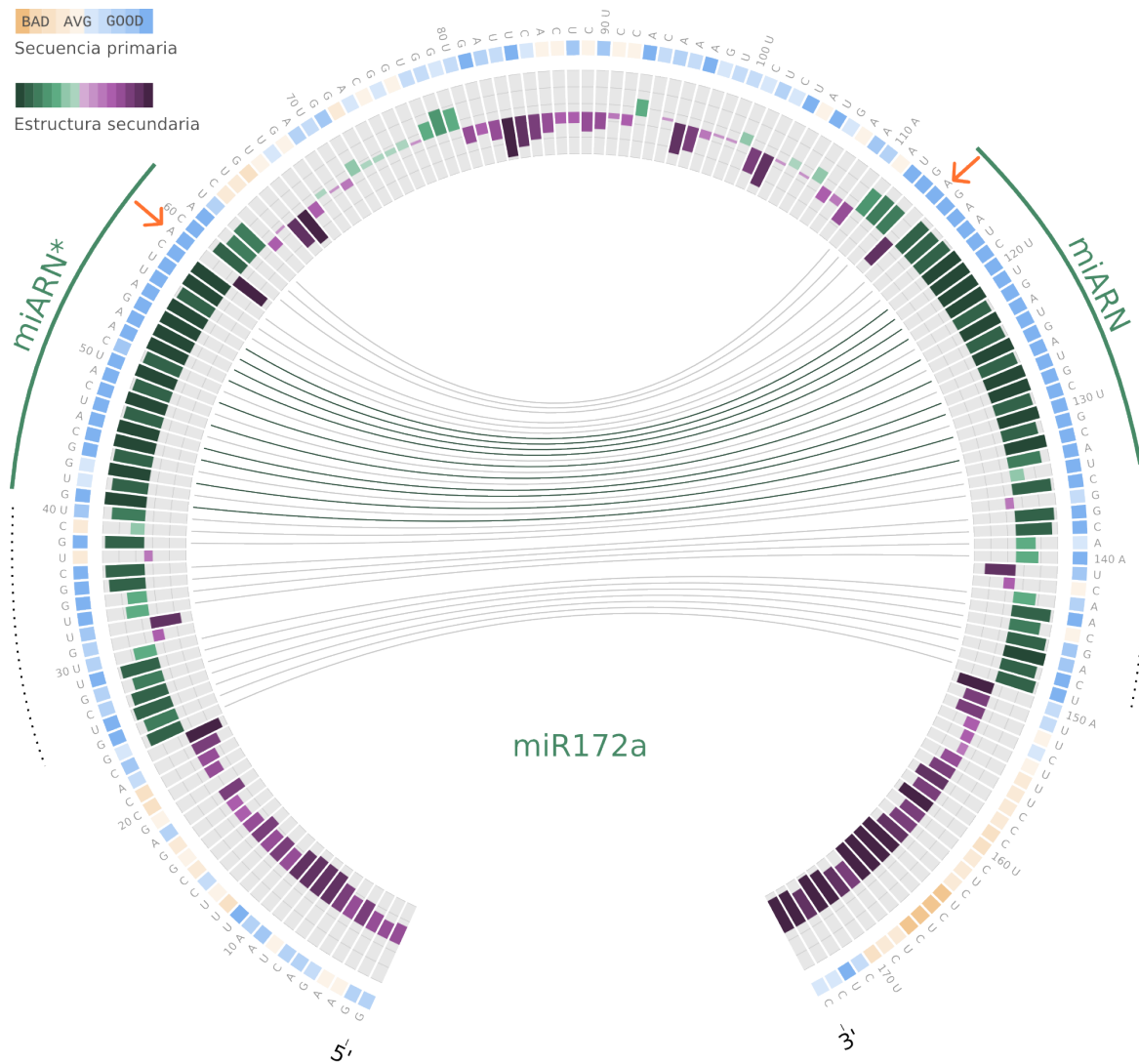


Figura 6.3: **Circos del miR172a.** En el anillo exterior se representa en colores el grado de conservación de la secuencia consenso a partir de los alineamientos en base a su secuencia primaria. Además, se muestran las bases del precursor de *A. thaliana*. Luego en el anillo interior se representa, mediante un histograma, la frecuencia de bases apareadas y desapareadas para cada posición del precursor en las distintas especies, donde una barra verde indica que la posición tiende a estar apareada formando una estructura secundaria, mientras que la barra violeta indica que la base tiende a estar desapareada. Con líneas de distinto espesor se muestra la interacción de las bases del precursor considerando la estructura secundaria. Fuera del anillo exterior se marca la secuencia del miARN y miARN\*.

### 6.2.2. Visualización por Circos de precursores que se procesan desde la base.

De acuerdo al análisis de las bibliotecas SPARE, en los precursores con un mecanismo corto de base a loop, un loop interno seguido por un tallo inferior de  $\sim 15$ nt especifica la posición del primer corte por debajo del miARN. El segundo corte procede a una distancia fija de  $\sim 21$  nt desde la posición del primer corte.

Para la mayoría de los precursores procesados con este mecanismo, pudimos observar el mismo patrón de conservación de estructura primaria y similares patrones de estructura secundaria que se muestra para el caso del precursor del miR172a (Figura 6.3) y del precursor del miR390a (Figura 6.4). En ambos casos los gráficos representan, ortólogos en las 30 especies utilizadas. También, se pudo observar que la región con mayor conservación coincide con el dúplex miARN/miARN\*, pero además, se puede ver una región conservada por debajo del dúplex, que coincide con el tallo inferior del precursor. Se puede observar que esta región está relativamente conservada a nivel de estructura primaria como secundaria.

Otra ventaja de esta forma de representar a los precursores con la conservación entre especies, es que visualmente se puede reconocer fácilmente las posiciones donde existen “mismatches” conservados. Por ejemplo, en el Circos del miR172, se puede observar que la primera base del miARN tiende a estar desapareada junto con la posición 19 del miARN\* (Figura 6.3, flecha naranja). Esta interacción es A-A y se mantiene sin variaciones en todas las especies analizadas (Figura 6.3). Un caso similar detectamos para el miR390a (Figura 6.4, flecha naranja).

Esto podría ser interesante estudiarlo en detalle, donde se podría generar un precursor mutante con esas bases en particular estén apareadas y luego ver si esto afecta o no el procesamiento del precursor.

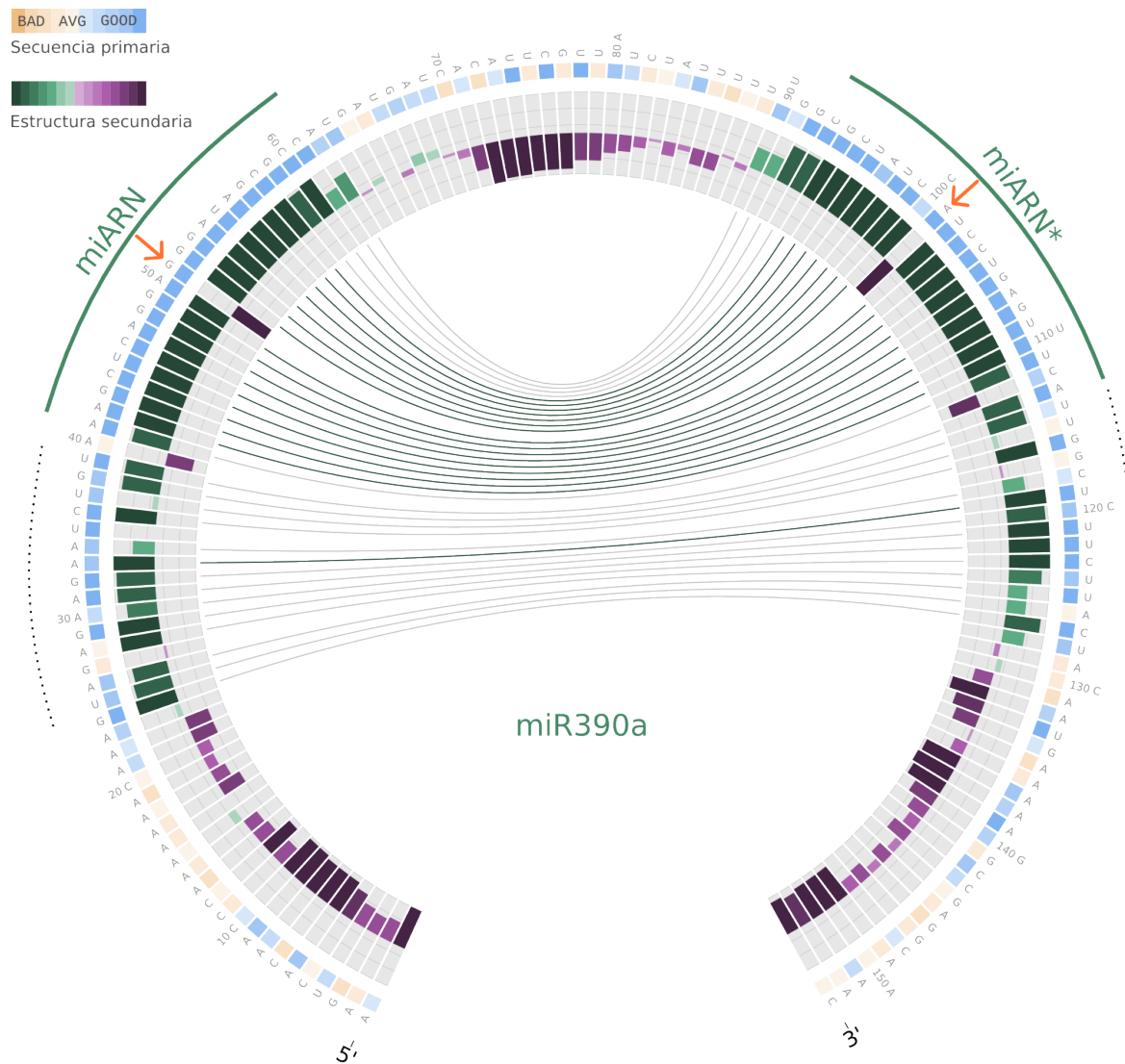


Figura 6.4: **Circos del miR390a.** Se puede observar que la región con mayor conservación coincide con el dúplex miARN/miARN\*. Además, se puede observar una región conservada por debajo del dúplex, que coincide con el tallo inferior.

### 6.2.3. Visualización por Circos de precursores que se procesan desde la base en forma secuencial.

En estos precursores, con un mecanismo secuencial de base a loop, el primer corte procede como en los precursores cortos de base a loop, pero luego son necesario dos cortes más para liberar el miARN, generando en el proceso niveles bajos de ARN pequeños adicionales [23]. Este es el caso de algunos miembros de la familias del miR169, que tiene en total 14 miembros siendo la familia más grande de *A. thaliana*. Un estudio en detalle de los precursores del miR169b/d/e/f/g mostró que los sitios de cortes del precursor estaban localizados 21 nt por debajo del lado proximal del dúplex miARN/miARN\*. Esta es la distancia esperada entre dos cortes de DCL1, sugiriendo que estos precursores eran procesados por más de dos cortes de la encima [23].

Además, la familia del miR394 también es procesada de la misma manera y comparten similares patrones de estructura secundaria, donde se ve el tallo inferior largo bien estructurado (Figura 6.5 B). Si observamos el patrón de conservación de secuencia primaria podemos ver que la región del dúplex miARN/miARN\* está muy conservada, pero además podemos observar que debajo del dúplex existe otra región conservada y estructurada, que coincide con el tallo inferior largo estructurado que presentan estos precursores (Figura 6.5). Es decir, que en estos precursores podemos también correlacionar su procesamiento con la conservación evolutiva.



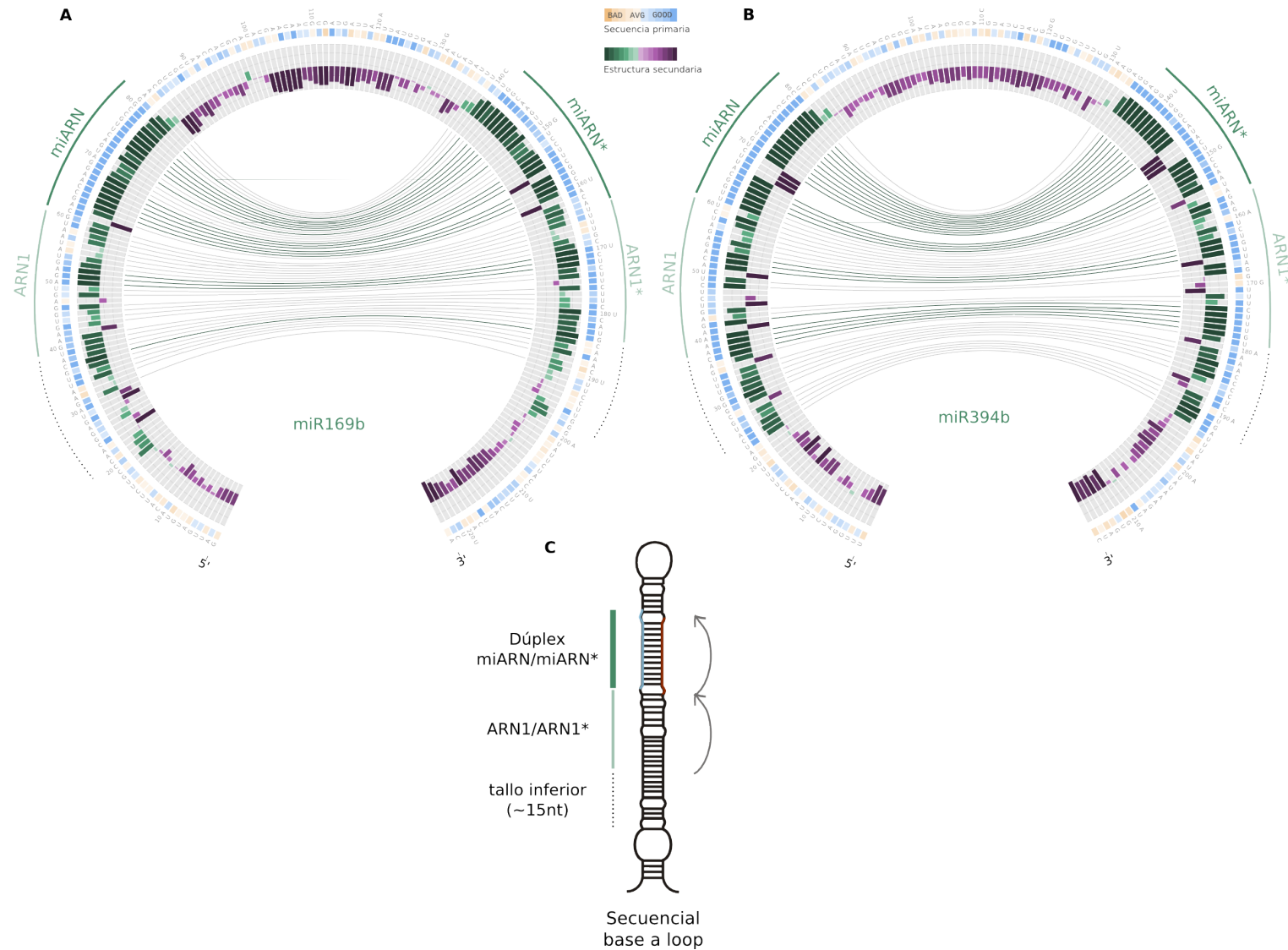


Figura 6.5: **Circos de precursores con mecanismos de procesamiento secuencial de base a loop** (A) Circos del precursor del miR169b. (B) Circos del precursor del miR394b. (C) Estructura de precursores procesados secuencialmente de base a loop. En (A) y (B) se puede observar que la región del dúplex miARN/miARN\* está muy conservada, pero además podemos observar que debajo del dúplex existe otra región conservada que coincide con el tallo inferior largo. Con líneas verde fuera del Circos se muestra la región que corresponde al miARN y al miARN\*. Con líneas verde claro fuera del Circos se muestran otros ARN pequeños generados a partir del primer y segundo corte por DCL1. Con líneas de puntos se muestra fuera del Circos la región que corresponde al tallo inferior de ~15nt.

#### **6.2.4. Visualización por Circos de precursores que se procesan desde el loop cortos.**

En los precursores con un mecanismo cortos de loop a base, el procesamiento es guiado por un tallo superior, y son necesarios dos cortes para liberar el miARN maduro. La región terminal de estos precursores tienen una largo conservado de  $\sim 42$ , donde incluye un loop pequeño [23] a diferencia de los precursores procesados de base a loop, donde esta región es variable en tamaño, como vimos en el capítulo anterior.

En este caso mostramos los Circos para el precursor del miR157a, que fue encontrado en 30 especies, y para el precursor del miR160a también encontrado en 30 especies (Figura 6.6). En los precursores que se procesan cortos de base a loop observamos que presentan una región conservada que corresponde al tallo superior además de la región que comprende al dúplex miARN/miARN\* (Figura 6.6). A diferencia de los precursores que se procesan de la base al loop, estos precursores no presentan el tallo inferior ni estructurado ni conservado y esto se puede observar en ambos casos; en el miR157a (Figura 6.6 A) y del miR160a (Figura 6.6 B).

#### **6.2.5. Visualización por Circos de precursores que se procesan desde el loop en forma secuencial.**

En los precursores con un mecanismo secuencial de loop a base, cuatro cortes secuenciales por DCL1 son los encargados de procesar los precursores de miARNs. En general, estos precursores muestran un tallo largo superior, del cual otros ARNs pequeños son generados [4, 21, 23].

En este caso se muestra a la familia del miR319 y la familia del miR159, y en particular mostramos al miR319a (Figura 6.7 A) y al miR159b (Figura 6.7 B). Observamos que la región que coincide con el tallo superior está conservada y además el tallo superior es estructurado. Además observamos que los ARNs pequeños, que se originan del procesamiento de estos precursores, están conservados en la estructura del precursor (Figura 6.7).

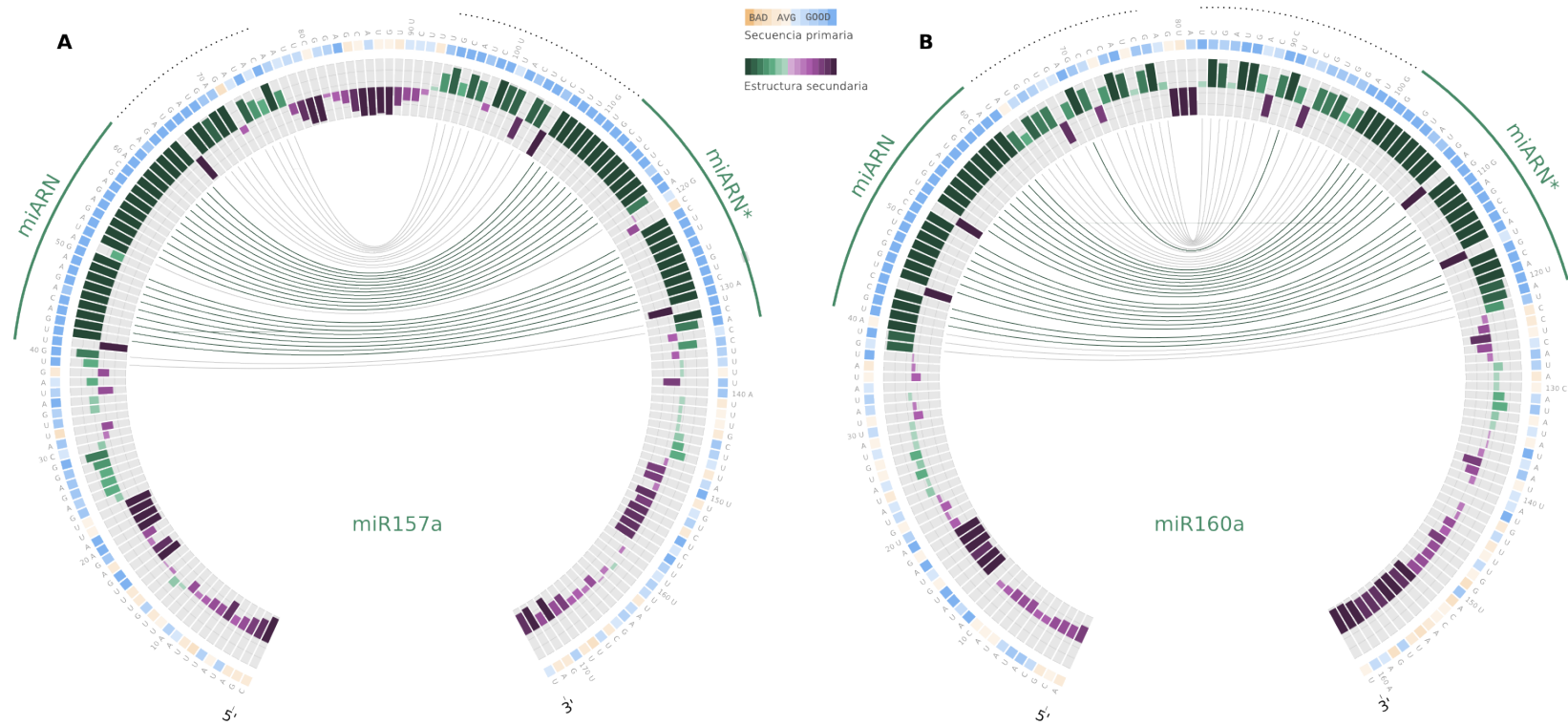


Figura 6.6: **Circos de precursores con mecanismos de procesamiento cortos de loop a base** (A) Circos del precursor del miR157a. (B) Circos del precursor del miR160a. (C) Estructura de precursores procesados cortos de loop a base. En (A) y (B) se puede observar que los otros ARNs pequeños están conservados de la misma manera que el miARN y el miARN\*. Con líneas verde fuera del Circos se muestra la región que corresponde al miARN y al miARN\*. Con líneas de puntos se muestra fuera del Circos la región que corresponde al tallo superior.

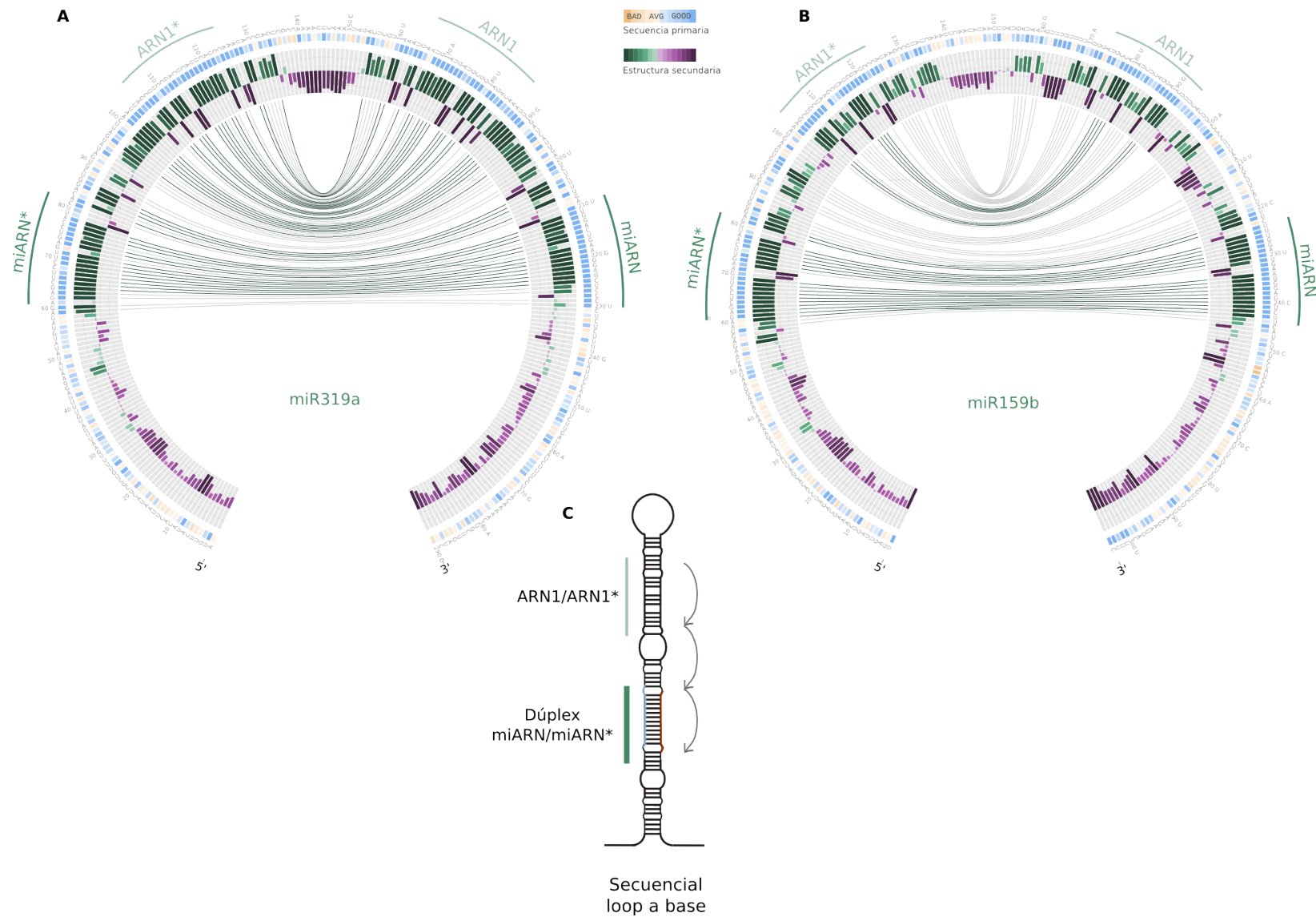


Figura 6.7: **Circos de precursores con mecanismos de procesamiento secuencial de loop a base** (A) Circos del precursor del miR319a. (B) Circos del precursor del miR159b. (C) Estructura de precursores procesados secuenciales de loop a base. En (A) y (B) se puede observar que los otros ARNs pequeños están conservados de la misma manera que el miARN y el miARN\*. Con líneas verde fuera del Circos se muestra la región que corresponde al miARN y al miARN\*. Con líneas verde claro fuera del Circos se muestran otros ARN pequeños generados a partir del primer y segundo corte por DCL1.

### **6.2.6. El tamaño del loop es muy variable en precursores que se procesan desde abajo.**

Una vez que hemos identificado los distintos precursores provenientes de estas 30 especies, decidimos hacer estudios adicionales sobre las dominios de los mismos. Para cada precursor en distintas especies, realizamos una búsqueda de motivos conservados con la herramienta MEME [87] (Figura 6.8 y 6.9). Hicimos esto para poder ver de manera visual cuáles eran los patrones comunes entre las distintas especies y donde se localizaban dentro del precursor. La búsqueda se hizo sobre elementos conservados de 20-23 nucleótidos y los parámetros utilizados se detallan en Materiales y Métodos (sección 3.6).

De acuerdo a la literatura, esperabamos que los motivos más conservados dentro del precursor sean el miARN y el miARN\*. Si bien las secuencias que encontramos como las conservadas se superponían tanto con el miARN y el miARN\*, en ocasiones observamos que estaban desfasada por uno o dos nucleótidos. Nuestra interpretación es que es posible que el sitio de corte de DCL1, que se extiende por fuera de la secuencia del miARN y el miARN\* tenga ciertos requisitos de estructura primaria y secundaria y de ahí su conservación.

Cuando visualizamos la localización de las secuencias conservadas en el contexto de los precursores que se procesan desde el loop observamos que las características del mismo estaban conservadas en las distintas especies. Vemos que el precursor del miR160a el tamaño de la región que comprende al tallo superior y al loop está conservada en tamaño y varía entre 35-40 nt (Figura 6.8).

Sin embargo cuando visualizamos los precursores que se procesan desde la base, pudimos observar que el tamaño de la región que comprende al tallo superior y al loop es muy variado en un mismo precursor en distintas especies. Para el miR172a esta región puede variar entre 40nt hasta 115nt (Figura 6.9). Interpretamos que esta variación se debe a que los determinantes estructurales para su procesamiento están en la base del precursor, mientras que la estructura del loop es proclive a variaciones sin afectar el procesamiento del mismo.

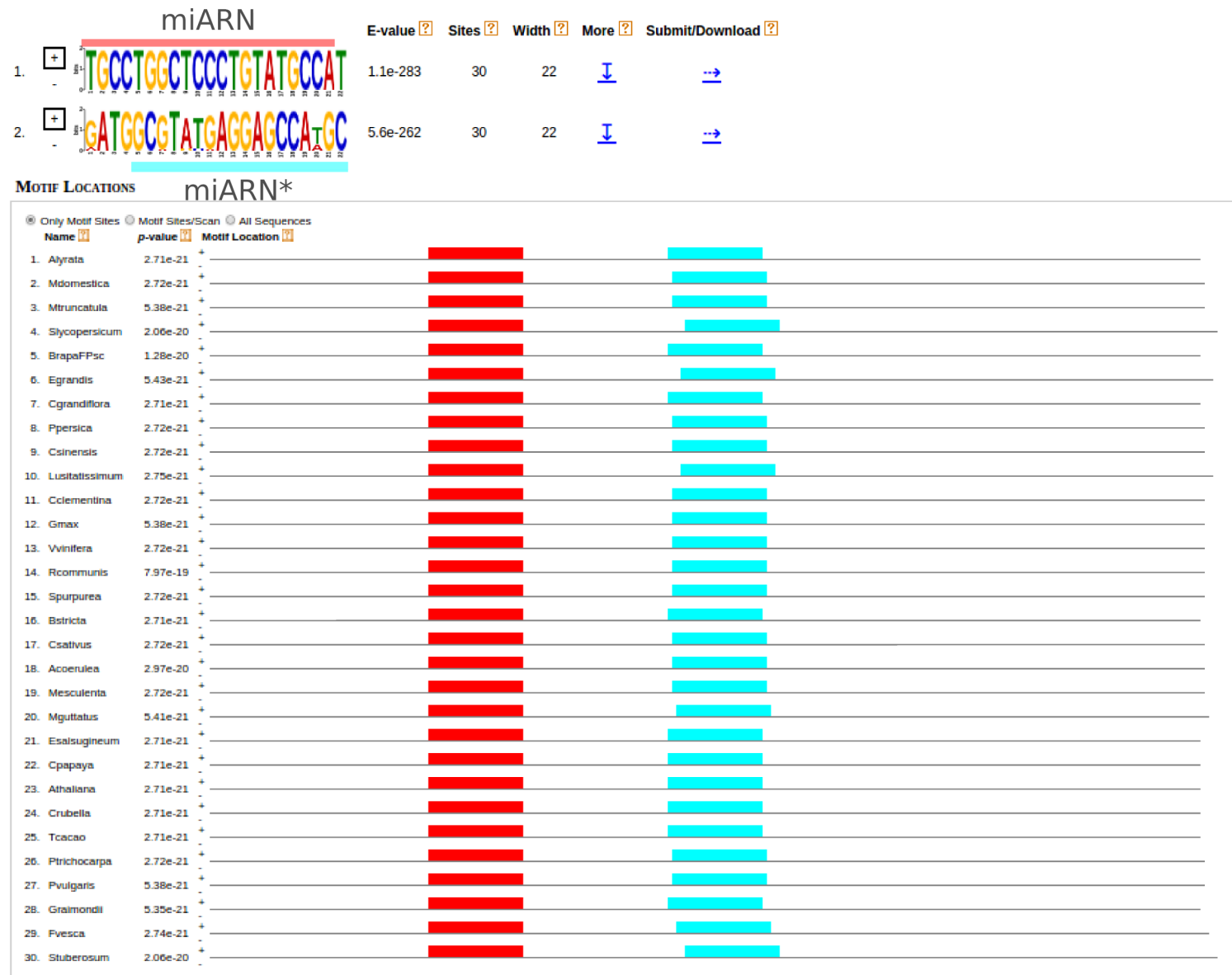


Figura 6.8: **Variación del loop terminal en precursores miR160** Se muestra el MEME donde se pueden observar los motivos conservados dentro de los precursores del miR160a en distintas especies. En color rojo el primer motivo conservado comprende parte del miR160 (logo 1). El segundo motivo más conservado comprende parte del miR160\* (logo 2). Se puede observar que la el tamaño de la región que comprende al tallo superior y al loop no varía en distintas especies.



Figura 6.9: **Variacion del loop terminal en precursores miR172a** Se muestra el MEME donde se pueden observar los motivos conservados dentro de los precursores del miR172a en distintas especies. En color rojo el primer motivo conservado comprende parte del miR172 (logo 1). El segundo motivo más conservado comprende parte del miR172\* (logo 2). Se puede observar que la el tamaño de la región que comprende al tallo superior y al loop es muy variado en distintas especies.

### 6.2.7. Procesamiento mixto de miembros de la familia del miR170/miR171.

En general, se considera que diferentes miembros de una misma familia de miARNs comparten la misma vía de biogénesis ya que se cree que las familias de miARNs se expanden por eventos de duplicación de un gen ancestral [6]. Sin embargo, ciertos el procesamiento de miembros de ciertas familias pueden variar de uno a otro [23]. Esto sucede, por ejemplo, en la familia del miR171 donde en *A. thaliana* existen tres miembros. El precursor del miR171a es procesado de base a loop, mientras que los precursores del miR171b y miR171c son procesados de loop a base.

Nos pusimos a analizar esta familia en detalle para ver si los patrones de conservación y estructura secundaria eran similares o diferentes (Figura 6.10). Observamos que el miR171a, que es procesado corto de base a loop, tiene un patrón de conservación similar a los precursores procesados de base a loop, donde el tallo inferior es estructurado y conservado, como era de esperarse (Figura 6.10 B). Por el contrario, el miR171c, que es procesado corto de loop a base, muestra conservación en el tallo superior, que además es estructurado, y no así el tallo inferior (Figura 6.10 A).

Esta familia mixta, presenta dos patrones claros y diferentes de conservación de secuencia primaria y de estructura secundaria. Demostrando que este análisis es sensible a la selección de ortólogos para miembros de una misma familia. Donde para el Circos del precursor del miR171a se incluyen 24 especies y para el Circos del precursor del miR171c se incluyen las 30 especies estudiadas.



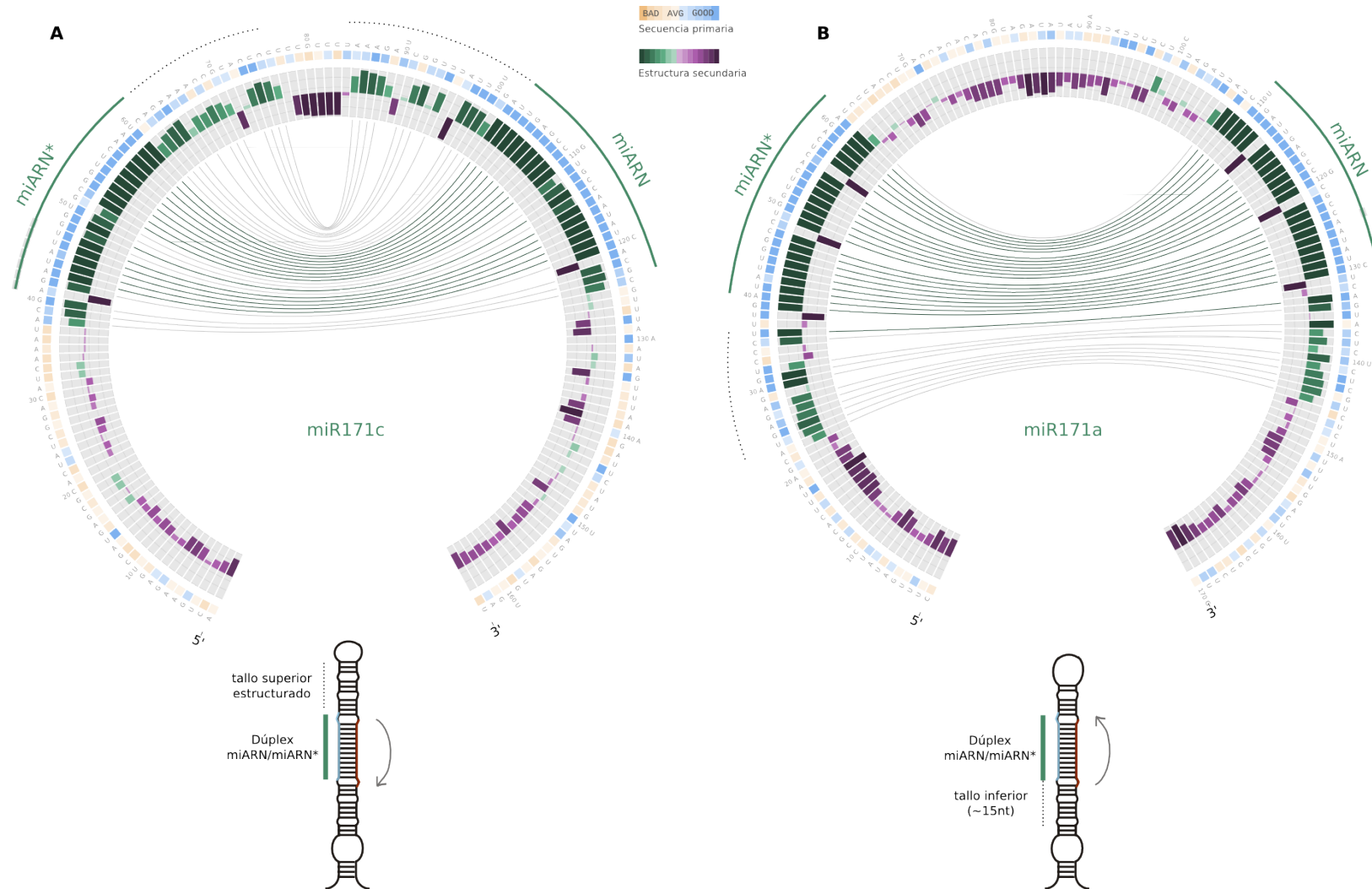


Figura 6.10: **Patrón de conservación de la familia con procesamiento mixto del miR170/miR171** (A) Circos del precursor del miR171c que es procesado de loop a base. (B) Circos del precursor del miR171a que es procesado de base a loop. Se puede observar que el miR171c, que es procesado corto de loop a base, muestra conservación en el tallo superior. Por el contrario el miR171a, que es procesado corto de base a loop, el tallo inferior es estructurado y conservado. Con líneas de puntos se muestra fuera del Circos la región que corresponde en (A) al tallo superior y en (B) al tallo inferior de ~15nt. Además, se muestra un esquema del mecanismo de procesamiento correspondiente a cada precursor.

### 6.2.8. Visualización por Circos de precursores con mutaciones puntuales que afectan el procesamiento de miARNs en plantas.

Se ha demostrado recientemente que un polimorfismo de origen natural que ocurre en el gen del miR164a afecta a la forma de la hoja y la arquitectura del vástago en *A. thaliana*, debido a una mutación presente en ecotipo C24, con los efectos de ser modificados por loci adicionales en el genoma [140]. Una sustitución única en un par de bases en la secuencia complementaria del miARN altera la estabilidad predicha del dúplex miARN/miARN\*. Se reduce con ello, en gran medida, la acumulación del miARN maduro, probablemente porque interfiere en el procesamiento del precursor. Además, se demostró en ese mismo artículo que no es una rara excepción y que las cepas naturales de *A. thaliana* albergan decenas de polimorfismos similares que afectan al procesamiento de una amplia gama de precursores de miARNs.

A nivel de secuencia los alelos de Col-0 y C24 son diferentes sólo por un par de polimorfismos simples. Uno de ellos afecta a una C en el miARN\* que aparea con una G en la posición dos del miARN (nos referimos a esta posición como \*2) (Figura 6.11 A).

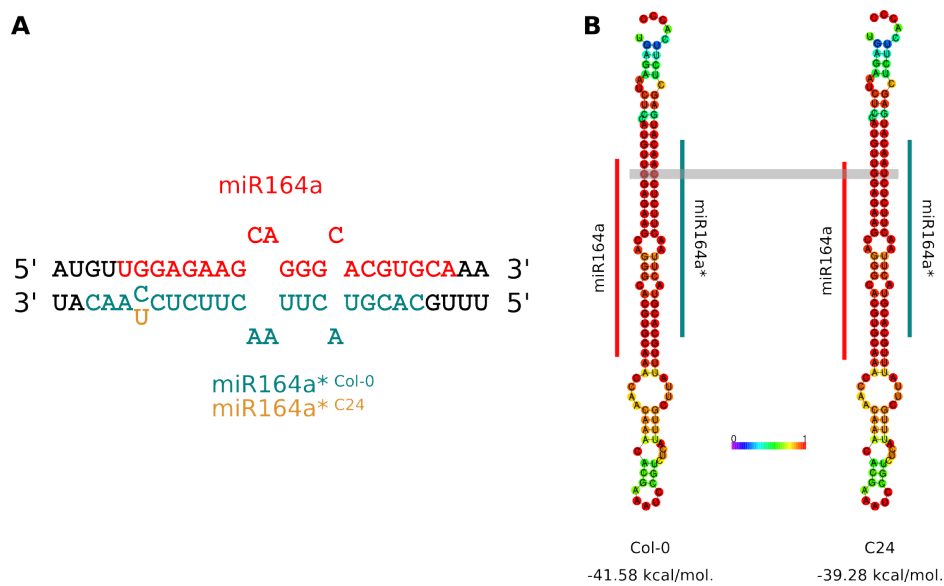


Figura 6.11: **Dúplex del miR164a/miR164a\* y estructura secundaria del precursor.** (A). Se muestra el apareamiento de bases en el precursor del miR164a. El miARN se muestra en rojo y el miARN\* en azul. En la posición \*2, el alelo alternativo C24 se muestra en naranja. (B). Se muestra la estructura secundaria predicha del precursor del miR164a. La barra gris horizontal indica el polimorfismo, un G:C en Col-0 y un G:U en C24. Los colores indican probabilidad de pares de bases.

Se puede ver que el remplazo de una G:C en el miR164a<sup>Col-0</sup> por un par no canónico G:U en el miR164a<sup>C24</sup>, incrementa ligeramente la energía libre del plegado del precursor que cambia de -41.58 kcal/mol a -39.28 kcal/mol. Pero se observa que no se modifica la estructura secundaria del miR164a si aceptamos que el par canónico G:U se forme en esta región (Figura 6.11 B).

Realizamos el Circos del miR164a en distintas especies para ver como era la conservación de la base en la posición 2\* que hace que el precursor no pueda ser procesado de manera correcta. Lo que pudimos observar es que dicha posición del miR164a\* está conservada en dicotiledóneas y se mantiene sin cambios en todas las especies (Figura 6.12). Además esa base está siempre apareada con la base correspondiente del miARN maduro, formando un G:C (Figura 6.12).

También se puede observar que en la secuencia del miR164a\* existen mutaciones en distintas bases en distintas especies, pero no en la base que estamos estudiando. Esto sugiere que esa base en particular es importante para la estabilidad del precursor y su buen procesamiento.

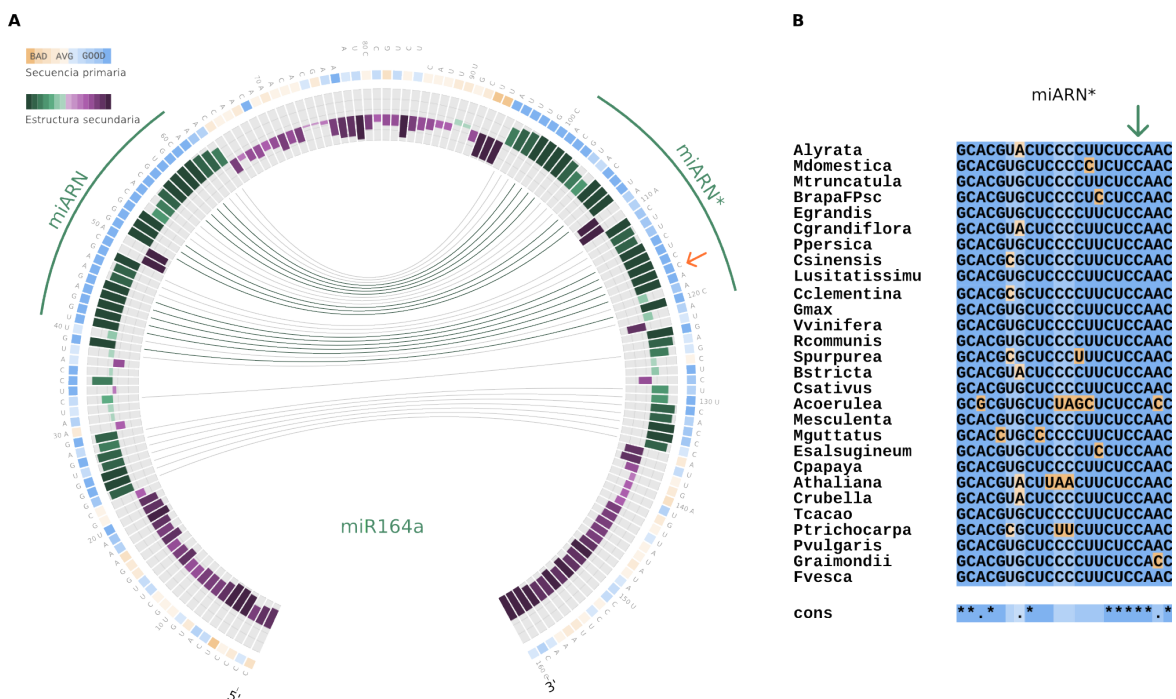


Figura 6.12: **Circos del miR164a.** (A) Se representa el Circos del miR164a. (B) Se representa El alineamiento del miR164a en distintas especies. Con una flecha naranja se destaca la posición \*2 a estudiar, en el Circos y en el alineamiento.

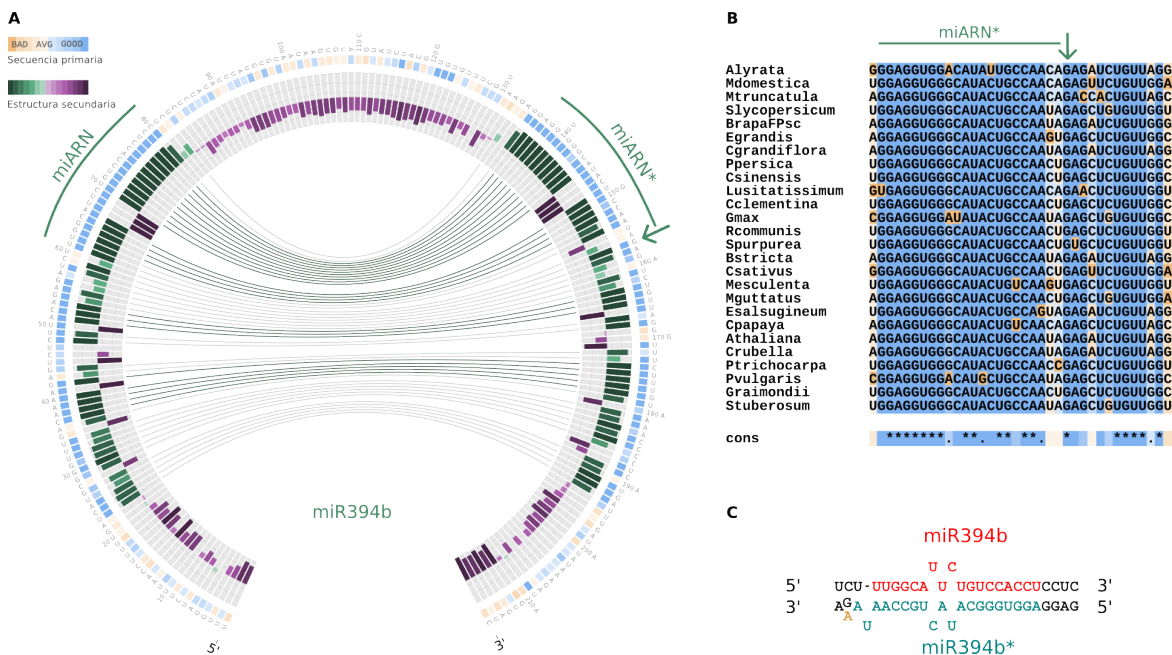


Figura 6.13: **Patrón de conservación del precursor del miR394b.** (A) Se muestra el Circos del miR394b. (B) Se muestra el alineamiento del miR394b en distintas especies. Con una flecha naranja se destaca la posición a estudiar, en el Circos y en el alineamiento. (C) Apareamiento de bases en el precursor del miR394b. El miARN se muestra en rojo y el miARN\* en azul. En naranja se muestra la posición de la mutación G por A en el tallo inferior del precursor del mir394b-1.

En otro trabajo, se estudió un alelo mutante del miR394b, llamado mir394b-1, que tiene un “mismatch” en el tallo inferior del precursor del miR394b [71]. La mutante incrementa dramáticamente la terminación del meristema apical. En la figura 6.13 C, se muestra el apareamiento de bases en el precursor del mir394b y mir394b-1, donde el miARN se muestra en rojo y el miARN\* en azul. Y en naranja se muestra la posición de la mutación del precursor del mir394b-1 donde la G es reemplazada por una A con respecto al mir394b. En este caso se mostró que los niveles de maduro en el precursor del mir394b-1 se ven fuertemente disminuidos aunque no completamente ausentes [71].

Para varios precursores de plantas hemos visto anteriormente que esa región que corresponde al tallo inferior es crucial para su procesamiento. En este caso mostramos el Circos correspondiente al miR394b, para poder visualizar que es lo que sucede con esta mutación puntual fuera del dúplex miARN/miARN\* (Figura 6.13). Vemos, como en el caso del miR164a, esta mutación que afecta al procesamiento del precursor y a la acumulación del miARN maduro, está conservada en todas las especies estudiadas. Esto sugiere que esta base es importante para el procesamiento y que no sólo mutaciones de una base dentro del

dúplex miARN/miARN\* pueden afectar el procesamiento, sino que mutaciones simples en el precursor puede afectar el reconocimiento de DCL1.

La metodología empleada y la visualización por medio de los Circos permite deducir cuales son las bases o apareamientos conservados en distintas especies. Pensamos además que estas bases son importantes por se para el procesamiento del precursor y la liberación del miARN maduro , ya que puede observarse una correlacion entre la conservación y las mutaciones puntuales descriptas. Esta información también podría ser utilizada para ayudar en el diseño de miARNs artificiales en distintas especies y aumentar su eficiencia.

### **6.2.9. Visualización por Circos de precursores de plantas conservados en monocotiledóneas.**

Se conocen muchas familias de precursores conservados en plantas, algunos de ellos conservados desde musgos hasta dicotiledóneas [13, 14, 38]. Si bien la mayoría de los precursores de plantas utilizados hasta esta parte del trabajo están conservados en dicotiledóneas y monocotiledóneas, muchos fueron evolucionando encontrandose nuevos miembros de algunas familias en especies más recientes y perdiendose otros. Es por esto que separamos a los precursores de plantas en dicotiledóneas y monocotiledóneas, y en esta parte del trabajo nos enfocamos en estas últimas. Las especies utilizadas de monocotiledóneas se pueden ver en la Figura 3.3 de Materiales y Métodos.

Existe una variante del miR396 que fue detectada por primera vez en arroz [135] y estudios posteriores indicaron que la misma es especifica de monocotiledóneas [136, 166]. El sitio complementario al miR396 de los GRFs (genes blanco del miR396) en dicotiledóneas posee un nucleótido, que en el apareamiento entre el miARN y el ARNm genera un bulge. Es muy interesante que la G extra (entre la posición 7-8 del miARN) presente en la variante de monocotiledóneas elimina este bulge en el apareamiento con el sitio blanco de los GRFs.

Realizamos el gráfico del Circos del miR396e de Arroz, donde pudimos detectar ortólogos en todas las plantas monocotiledóneas estudiadas en nuestro trabajo (Figura 6.14). Observamos que el nucleótido extra, que le da identidad a la variante de monocotiledóneas, está conservado en todas las especies. Se observa que a nivel de secuencia primaria el precursor está conservado, pero el patrón de conservación no coincide exactamente con los patrones de conservación vistos anteriormente para los precursores que se procesan desde la base como los que se procesan desde el loop. En particular se observa conservación de secuencia primaria y secundaria por arriba y por debajo del dúplex (Figura 6.14).

Este análisis en monocotiledóneas, demuestra que el enfoque bioinformático para el estudio de la evolución de precursores de miARNs en plantas también puede ser utilizado para encontrar determinantes estructurales en un grupo de especies cercanas evolutivamente. Además, el mismo permite poder comparar patrones de evolución en distintos grupos de especies.

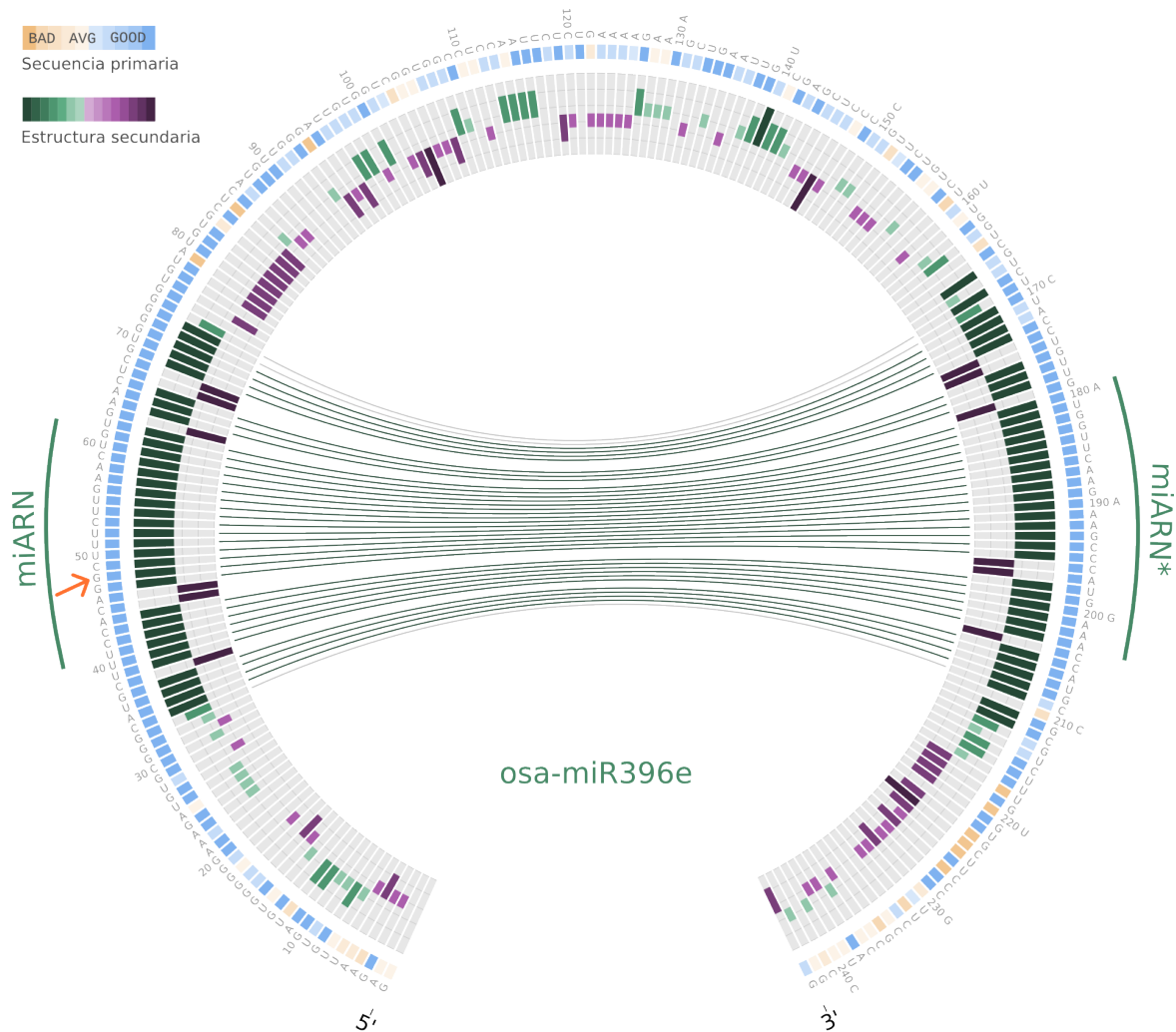


Figura 6.14: **Circos del miR396e de monocotiledóneas** Se puede observar que el nucleótido extra, que le da identidad a la variante de monocotiledóneas, está conservado en todas las especies. Este nucleótido extra que se marca con una flecha naranja.



### 6.2.10. Estrategia bioinformática para estudiar precursores de miARNs en animales.

En animales existe un gran número de miARNs que están conservados en un gran rango de especies, incluso se han encontrado algunos miARNs conservados en humanos hasta gusanos [110]. Si bien los miARNs en plantas y en animales comparten componentes básicos, su biogénesis y su modo de acción es diferente. Una de las grandes diferencias con las plantas radica en que en animales el procesamiento de los precursores ocurre en dos pasos separados en tiempo y espacio (Figura 1.2), mientras que en plantas ocurre íntegramente en el núcleo (Figura 1.3).

Además de las diferencias en la biogénesis, otra de las grandes diferencias entre plantas y animales consiste en las estructuras de los precursores de miARNs. Los precursores de animales se caracterizan por presentar una estructura relativamente homogénea, que consta de un tallo imperfecto de una longitud de ~65 nt, flanqueado por dos regiones desapareadas. Estas dos regiones son un bucle terminal en un extremo y ARNsh en el otro. En animales, la especificidad en la selección de la secuencia del miARN dentro del precursor está dada por el primer corte realizado por el complejo microprocesador. Por otro lado, los precursores de miARNs de plantas presentan tamaños que varían desde los 50 a 900 nt, lo que se traduce en un amplio abanico de estructuras diferentes.

Para esta parte del trabajo de Tesis, utilizamos genomas de especies de *Metazoa* entre ellos humano, mono, sapo, vaca y pez (Tabla 6.1) y estudiamos que sucede con los precursores de miARNs conservados en animales. Para esto partimos de los precursores de humanos definidos en miRBase y realizamos la búsqueda de ortólogos en otras especies al igual que lo hicimos en plantas.

En la Figura 6.15 se muestra el Circos del precursor del miARN hsa-let-7a-1, un miembro de la familia let-7. El precursor del miARN let-7 identificado en un estudio en *C. elegans* fue uno de los primeros miARNs en ser descubiertos. Actualmente ha sido predicho y validado en un gran número de especies.

En un artículo publicado en Nat. Struct. Mol. Biol., se encontró una ribonucleoproteína (hnRNP A1) que actúa como regulador negativo de let-7-a [97]. HnRNP A1 se une al loop terminal del precursor de let-7-a e inhibe su procesamiento por Drosha. En el Circos se puede observar que este loop terminal está conservado en las especies utilizadas (6.15).

La variación de los precursores de miARNs de plantas se debe en gran parte a la variación en el tamaño del loop. Observamos el tamaño del loop terminal en precursores de plantas

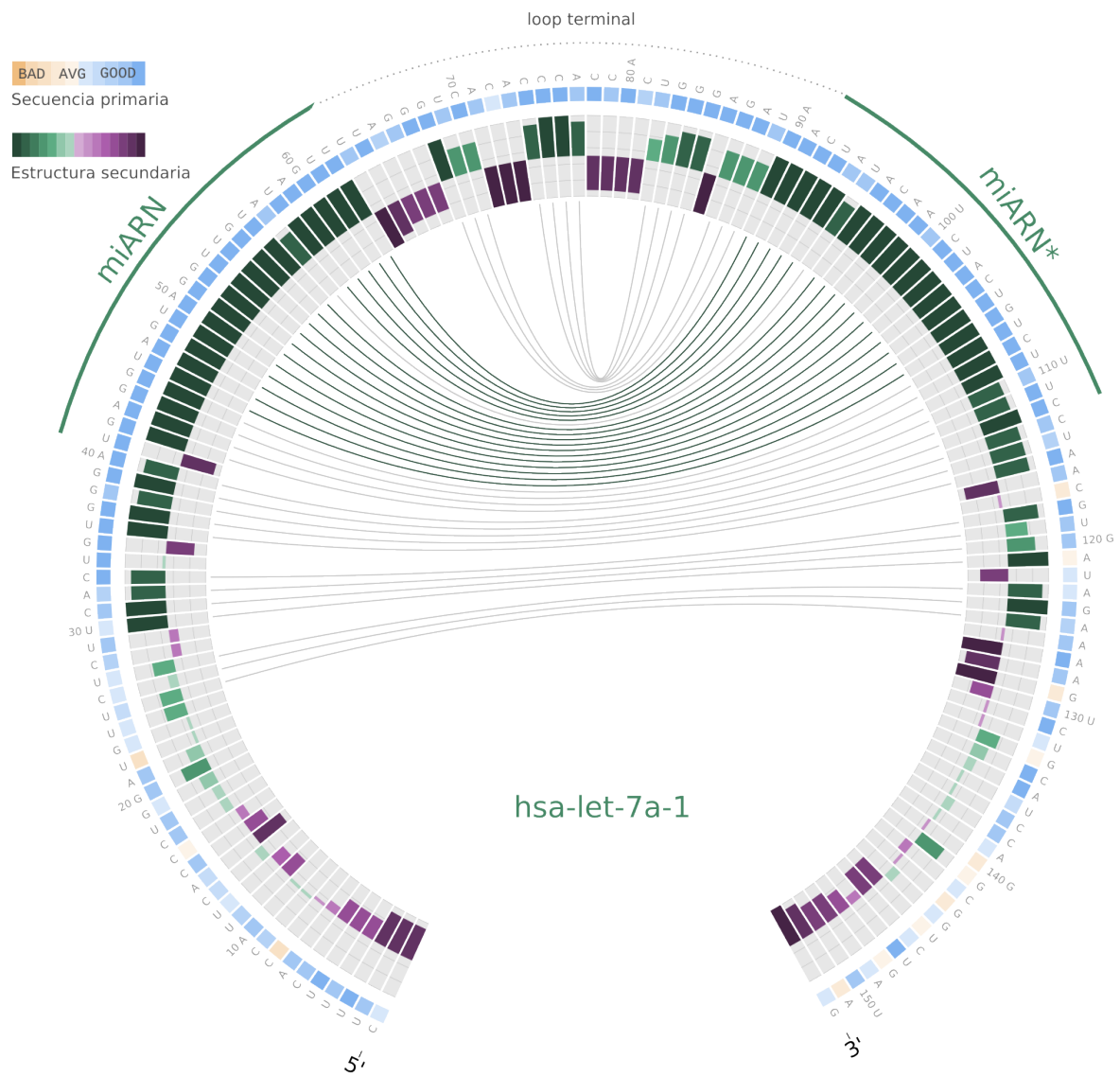


Figura 6.15: **Patrón de conservación del precursor del precursor del miARN.**  
Se puede observar que loop terminal (marcado en línea de puntos) está conservado en las especies utilizadas.



Tabla 6.1: Especies de *Metazoa* utilizadas

<b>Animales</b>
Bos taurus
Canis familiaris
Equus caballus
Gallus gallus
Gorilla gorilla
Homo sapiens
Macaca mulatta
Monodelphis domestica
Mus musculus
Ornithorhynchus anatinus
Petromyzon marinus
Sus scrofa
Xenopus tropicalis

procesados cortos de base a loop o de loop a base y lo comparamos con el tamaño del loop terminal en precursores conservados en animales. Se puede notar que el tamaño del loop en los precursores de plantas cortos de base a loop son más variables además de más grande. En cambio el tamaño del loop terminal de los precursores de humano son pequeños y similares al tamaño del loop terminal de precursores de plantas loop a base.

En animales, la estructura secundaria del loop terminal de precursores de miARNs, es irrelevante durante el procesamiento e inclusive se demostró que la presencia del mismo es innecesaria para el procesamiento por DROSHA [62]. Sin embargo, la estructura del loop es importante para que los pre-miARNs sean exportados del núcleo al citoplasma, donde DICER produce el segundo corte que libera a los miARNs en animales. El mecanismo de procesamiento en animales es similar al mecanismo de procesamiento de los precursores de plantas cortos de base a loop. Sin embargo, podemos observar que estructuralmente los precursores en animales comparten patrones similares a los precursores de plantas cortos de loop a base y también a los cortos de la base al loop (Figura 6.16) . Esto se debería a que en animales existen determinantes para guiar a DROSHA (por debajo del miARN) y otros determinantes necesarios para que se exporte el precursor al citoplasma, donde DCL1 produce el segundo corte.

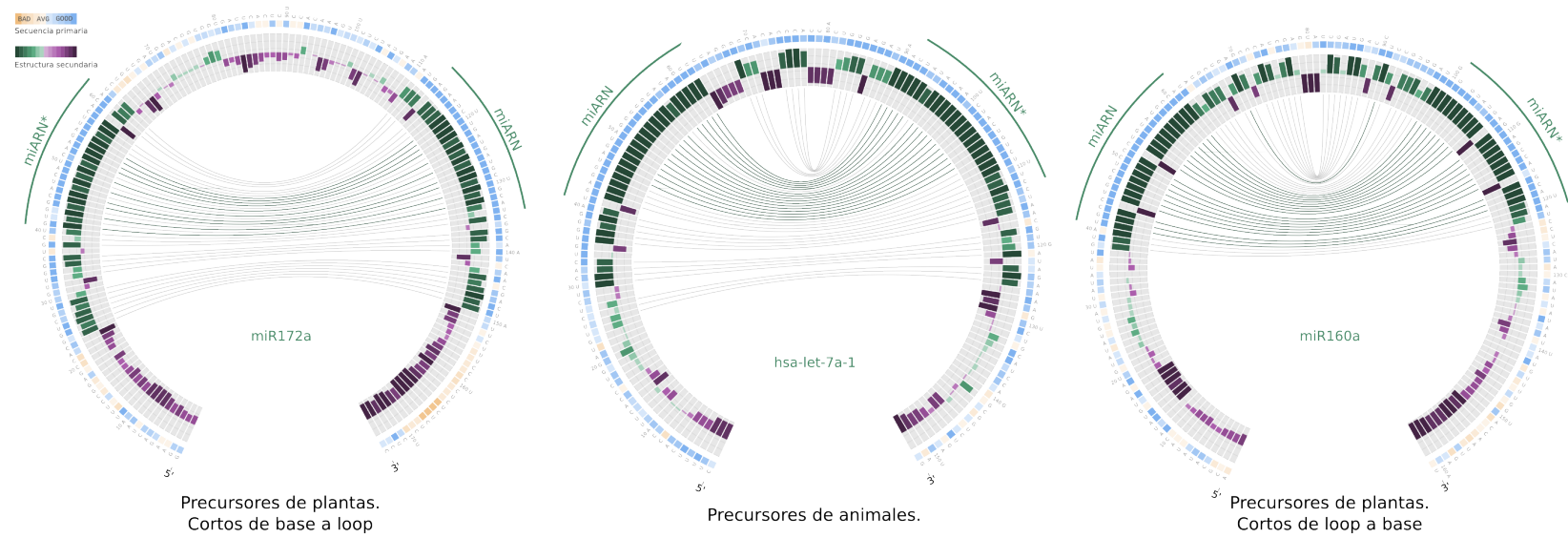


Figura 6.16: **Circos de animales vs plantas.** A la izquierda el Circos del miR172a, un precursor conservado en plantas con mecanismo corto de base a loop. En el medio el Circos del precursor de miARN de humano let-7a-1. A la derecha el Circos del miR160a, un precursor conservado en plantas con mecanismo corto de loop a base.

### 6.3. Conclusiones.

En esta parte del proyecto de Tesis presentamos un enfoque bioinformático para el estudio de la evolución y biogénesis de miARNs en plantas. Para esto desarrollamos una implementación gráfica para visualizar de manera simple los precursores de miARNs en distintas especies de plantas. Y fueron utilizados para caracterizar la evolución de precursores de miARNs en plantas con distintos mecanismos de procesamiento. En el Anexo (Figuras 9.1,9.2,9.3,9.4 y 9.5) se muestran los Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas.

Esta metodología y visualización con Circos, la utilizamos para comparar precursores de miARNs de distintas familias y además en precursores de miARNs de distintos reinos. En particular, estudiamos precursores con mutaciones puntuales que afectan al procesamiento de miARNs en plantas. Y vimos, que la información obtenida de estos estudios, podría ser utilizada para ayudar en el diseño de miARNs artificiales en distintas especies y aumentar su eficiencia.

Si bien la biogénesis y modo de acción de precursores de animales es diferente a la de los precursores de plantas, pudimos utilizar este mismo enfoque para estudiar precursores de miARNs en animales. Y pudimos observar que estructuralmente los precursores en animales comparten patrones similares a los precursores de plantas cortos de loop a base y también a los cortos de la base al loop. Esto podría ser debido a que en animales existen determinantes para guiar a DROSHA (por debajo del miARN) y otros determinantes para que el exporte al citoplasma (por encima del miARN).

# Conclusiones finales.

## 8.1. Aplicaciones bioinformáticas para el estudio de interacciones miARN-gen blanco.

En cuanto a la primera parte de la Tesis, diseñamos una estrategia para identificar genes blanco regulados por miARNs en plantas, basada en la conservación evolutiva del par miARN-gen blanco. Los estudios previos sobre la identificación de genes blanco de miARN se basaban en que la interacción entre el ARN pequeño y el ARNm tenga ciertos requerimientos en cuanto al número de “mismatches”, su posición relativa y la energía de interacción [17, 30, 85]. En algunos casos, se analizaba la conservación del par miARN/miARN\*, fundamentalmente entre *Arabidopsis* y arroz [149], como elemento confirmatorio. En este caso, nosotros hemos basado la búsqueda en la conservación evolutiva como elemento central, haciendo la búsqueda en más de 40 especies, aunque este número es fácilmente ampliable en la medida que haya disponible información genómica de otras especies. Si bien nosotros nos hemos enfocado en transcritos validados experimentalmente, también podría utilizarse como elementos de búsquedas en genomas secuenciados y predicciones de genes teóricos.

El enfoque desarrollado requiere que la interacción miARN-gen blanco, pueda ocurrir en el contexto de un conjunto mínimo de parámetros en diferentes especies. La secuencia del gen blanco en sí, no necesariamente tiene que estar conservada. Además, nuestro enfoque permite ajustar el número de especies requeridas como un filtro para realizar la búsqueda con diferentes sensibilidades y relaciones señal/ruido. Utilizando esta estrategia identificamos y validamos experimentalmente nuevos genes blanco en *A. thaliana*, a pesar de que este sistema ya había sido estudiado en detalles en distintos enfoques genómicos a gran escala ([2, 5, 53, 66, 112, 127]). Tres de los nuevos genes blanco validados tienen bulges. Parámetros empíricos usualmente le otorgan una gran penalidad a ellos, que puede llegar a ser el doble que un “mismatch” regular [66], sin embargo es probable que genes blancos con bulges asimétricos sean más frecuente de lo que se pensaba previamente en plantas. El enfoque

ofrece una estrategia alternativa a otras predicciones que se basan en parámetros empíricos del par miARN-gen blanco [5, 38, 47, 66]. Una ventaja de la estrategia presentada es que la interacciones miARN-gen blanco conservadas probablemente participen en procesos biológicos relevantes, ya que está en duda el rol biológico de la regulación por miARNs en interacciones de aparición reciente [14, 38].

## **8.2. Desarrollo de la herramienta comTAR.**

Como parte del estudio de las redes de miARNs y genes blanco de esta Tesis, desarrollamos una herramienta web denominada comTAR para predecir potenciales genes blanco regulados por miARNs en plantas. Esta herramienta, es intuitiva y flexible, y puede ser utilizada para buscar familias de potenciales genes blanco de miARNs en plantas e incluso para predecir potenciales genes blanco para nuevos ARN pequeños. La herramienta cuenta con datos pre-analizados para los miARNs conservados, pero el usuario puede hacer búsquedas con cualquier secuencia. Así por ejemplo, un miARN recientemente descubierto en monocotiledóneas puede ser usado para buscar blancos presentes en esas especies. En este caso sería de esperar que los blancos para este miARN estén conservados solamente en monocotiledóneas, por lo que búsqueda una búsqueda de potenciales blancos en general puede servir como validación general de la red regulatoria.

La búsqueda de blancos de miARNs también da como resultado la captura de múltiples interacciones en distintas especies. Así por ejemplo la búsqueda de genes blanco de miR319 arroja como resultados genes blanco con homología a TCP4 (At3g15030, TAG de Arabidopsis) en 31 especies. Si bien las interacciones entre el miR319 y TCP4 están relativamente conservadas, existen variaciones. Estas variaciones parecen estar localizadas en posiciones específicas. El análisis de estas variaciones es algo que podrá hacerse en el futuro y podría servir para mejorar el conocimiento sobre el apareamiento miARN gen blanco en general, así como ayudar a comprender la co-evolución de las secuencias de miARNs y sus genes blanco.

## **8.3. Determinantes mecánicos del procesamiento de miARNs en animales.**

En la segunda parte de esta Tesis, desarrollamos un enfoque genómico a gran escala para estudiar el procesamiento de miARNs en plantas y determinamos, de esta manera, el mecanismo de procesamiento de la mayoría de los miARNs de *A. thaliana* conservados evo-

lutivamente [23]. Encontramos que los miARNs en plantas pueden ser procesados por cuatro mecanismos, dependientes de la dirección secuencial de la maquinaria de procesamiento y del número de cortes requeridos para liberar el miARN [23]. Cuando los precursores de miARNs de plantas tratan de ser analizados en forma conjunta resultan ser muy heterogéneos y sin dominios comunes claros. Pudimos observar que cuando los precursores procesados con el mismo mecanismo son analizados como un grupo, entonces vemos que comparten determinantes estructurales claros. La co-existencia de distintos mecanismos cada uno con perfiles de procesamiento particulares explicaría el rango de tamaños y formas observados en precursores de miARNs en plantas [23].

Cuando los precursores se procesan desde la base hacia el loop vemos que existe un ARNdh de aproximadamente 15 bases por debajo de la región del miARN/miARN\*. Sin embargo, si el procesamiento comienza desde el loop, vemos que por arriba de la región del miARN/miARN\* existe una región estructurada doble hebra de aproximadamente 17 bases. La situación se hace aun más compleja con la posibilidad de múltiples cortes de DCL1, existiendo una o dos regiones estructuradas similares al miARN/miARN\* localizadas en forma continua.

#### **8.4. Desarrollo de una herramienta para el análisis de bibliotecas de SPARE incluyendo una interfaz gráfica.**

Además, desarrollamos una estrategia bioinformática para analizar las bibliotecas de SPARE y luego una herramienta para visualizar los resultados. Esto permite estudiar a los intermediarios de procesamiento en plantas silvestres y/o en plantas mutantes en proteínas de procesamiento o crecidas en distintas condiciones. Esta herramienta puede ser utilizada para determinar la dirección de procesamiento de precursores de miARNs y además para buscar diferencias en los cortes determinados por DCL1 en distintas condiciones o la precisión de la maquinaria en determinadas en mutantes.

### **8.5. Análisis de las estructuras de los precursores y su evolución. Visualización de información compleja por adaptación de una herramienta Circos.**

También, realizamos un enfoque bioinformático para el estudio de la evolución y biogénesis de miARNs en plantas, capturando precursores de miARN ortólogos de 40 especies diferentes. Para analizar la conservación de la estructura secundaria de los distintos precursores y poder comparar entre precursores de distinto procesamiento, diseñamos una implementación gráfica. Utilizando estos gráficos llamados Circos, estudiamos y caracterizamos la evolución de precursores de miARNs en plantas con distintos mecanismos de procesamiento. Además, mostramos que este enfoque puede ser utilizado también para estudiar precursores de miARNs en animales.

# Bibliografía.

- [1] Addo-Quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Curr. Biol.*, 18(10):758–762.
- [2] Addo-quaye, C., Eshoo, T. W., Bartel, D. P., and Axtell, M. J. (2009). Endogenous siRNA and microRNA targets identified by sequencing of the Arabidopsis degradome. *NIH Public Access*, 18(10):758–762.
- [3] Addo-Quaye, C., Miller, W., and Axtell, M. J. (2009a). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics*, 25(1):130–131.
- [4] Addo-Quaye, C., Snyder, J. A., Park, Y. B., Li, Y. F., Sunkar, R., and Axtell, M. J. (2009b). Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the *Physcomitrella patens* degradome. *RNA*, 15(12):2112–2121.
- [5] Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell*, 121(2):207–221.
- [6] Allen, E., Xie, Z., Gustafson, A. M., Sung, G. H., Spatafora, J. W., and Carrington, J. C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat. Genet.*, 36(12):1282–1290.
- [7] Allen, R. S., Li, J., Stahle, M. I., Dubroue, A., Gubler, F., and Millar, A. A. (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. *Proc. Natl. Acad. Sci. U.S.A.*, 104(41):16371–16376.
- [8] Altschup, S. F., Gish, W., Pennsylvania, T., and Park, U. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, pages 403–410.
- [9] Anders, G., Mackowiak, S. D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., Landthaler, M., and Dieterich, C. (2012). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, 40(Database issue):D180–186.
- [10] Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D. C. (2005). Cloning and characterization of micro-RNAs from moss. *The Plant journal : for cell and molecular biology*, 43(6):837–48.
- [11] Aukerman, M. J. and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell*, 15(11):2730–2741.



- [12] Axtell, M. J. (2008). Evolution of microRNAs and their targets: are all microRNAs biologically relevant? *Biochimica et biophysica acta*, 1779(11):725–34.
- [13] Axtell, M. J. and Bartel, D. P. (2005). Antiquity of microRNAs and their targets in land plants. *Plant Cell*, 17(6):1658–1673.
- [14] Axtell, M. J. and Bowman, J. L. (2008). Evolution of plant microRNAs and their targets. *Trends in Plant Science*, 13(7):343–349.
- [15] Axtell, M. J., Westholm, J. O., and Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, 12(4):221.
- [16] Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, 2:28–36.
- [17] Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2):215–233.
- [18] Bartel, D. P., Lee, R., and Feinbaum, R. (2004). MicroRNAs: Genomics , Biogenesis , Mechanism , and Function Genomics : The miRNA Genes. *Cell*, 116:281–297.
- [19] Baulcombe, D. (2004). RNA silencing in plants. *Nature*, 431(7006):356–363.
- [20] Bo, X. and Wang, S. (2005). TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics*, 21(8):1401–1402.
- [21] Bologna, N. G., Mateos, J. L., Bresso, E. G., and Palatnik, J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *The EMBO journal*, 28(23):3646–56.
- [22] Bologna, N. G., Schapire, A. L., and Palatnik, J. F. (2012). Processing of plant microrna precursors. *Briefings in Functional Genomics*.
- [23] Bologna, N. G., Schapire, A. L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B. C., and Palatnik, J. F. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome research*, 23(10):1675–89.
- [24] Bonnet, E., He, Y., Billiau, K., and Van de Peer, Y. (2010). TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, 26(12):1566–1568.
- [25] Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., and Zhu, J. K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell*, 123(7):1279–1291.
- [26] Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J. B., Crete, P., Chen, X., and Vaucheret, H. (2003). Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr. Biol.*, 13(10):843–848.
- [27] Bowman, J. L. (2004). Class III HD-Zip gene regulation, the golden fleece of ARGO-NAUTE activity? *Bioessays*, 26(9):938–942.

- [28] Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, 320(5880):1185–1190.
- [29] Brousse, C., Liu, Q., Beauclair, L., Deremetz, A., Axtell, M. J., and Bouche, N. (2014). A non-canonical plant microRNA target site. *Nucleic Acids Res.*, 42(8):5270–5279.
- [30] Carrington, J. C. and Ambros, V. (2003). Role of microRNAs in plant and animal development. *Science*, 301(5631):336–338.
- [31] Chapman, E. J. and Carrington, J. C. (2007). Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.*, 8(11):884–896.
- [32] Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J. R., and Belostotsky, D. A. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell*, 131(7):1340–1353.
- [33] Chen, H. M., Chen, L. T., Patel, K., Li, Y. H., Baulcombe, D. C., and Wu, S. H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 107(34):15269–15274.
- [34] Chen, X. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science*, 303(5666):2022–2025.
- [35] Chorostecki, U., Crosa, V. A., Lodeyro, A. F., Bologna, N. G., Martin, A. P., Carrillo, N., Schommer, C., and Palatnik, J. F. (2012). Identification of new microRNA-regulated genes by conserved targeting in plant species. *Nucleic Acids Research*.
- [36] Chorostecki, U. and Palatnik, J. F. (2014). comTAR: a web tool for the prediction and characterization of conserved microRNA targets in plants. *Bioinformatics (Oxford, England)*, 30(14):2066–7.
- [37] Cuperus, J. T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R. T., Takeda, A., Sullivan, C. M., Gilbert, S. D., Montgomery, T. A., and Carrington, J. C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat. Struct. Mol. Biol.*, 17(8):997–1003.
- [38] Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of mirna genes. *The Plant cell*, 23(2):431–442.
- [39] Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., and Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.*, 139(1):5–17.
- [40] Dai, X. and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, 39(Web Server issue):W155–159.
- [41] Davies, R. T., Goetz, D. H., Lasswell, J., Anderson, M. N., and Bartel, B. (1999). IAR3 Encodes an Auxin Conjugate Hydrolase from Arabidopsis. *Plant Cell*, 11(March):365–376.

- [42] Debernardi, J. M., Rodriguez, R. E., Mecchia, M. A., and Palatnik, J. F. (2012). Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions. *PLoS Genet*, 8(1):e1002419.
- [43] Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–235.
- [44] Ding, Y., Chan, C. Y., and Lawrence, C. E. (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, 32(Web Server issue):W135–141.
- [45] Dugas, D. V. and Bartel, B. (2008). Sucrose induction of Arabidopsis miR398 represses two Cu/Zn superoxide dismutases. *Plant Mol. Biol.*, 67(4):403–417.
- [46] Fabian, M. R., Sonenberg, N., and Filipowicz, W. (2010). Regulation of mRNA Translation and Stability by microRNAs. *Annual Review of Biochemistry*, 79(79):351–79.
- [47] Fahlgren, N. and Carrington, J. (2010). mirna target prediction in plants. In Meyers, B. C. and Green, P. J., editors, *Plant MicroRNAs*, volume 592 of *Methods in Molecular Biology*, pages 51–57. Humana Press.
- [48] Fahlgren, N., Jogdeo, S., Kasschau, K. D., Sullivan, C. M., Chapman, E. J., Laubinger, S., Smith, L. M., Dasenko, M., Givan, S. a., Weigel, D., and Carrington, J. C. (2010). MicroRNA gene evolution in Arabidopsis lyrata and Arabidopsis thaliana. *The Plant cell*, 22(4):1074–89.
- [49] Floyd, S. K. and Bowman, J. L. (2004). Gene regulation: ancient microRNA target sequences in plants. *Nature*, 428(6982):485–486.
- [50] Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19(1):92–105.
- [51] Gasciolli, V., Mallory, A. C., Bartel, D. P., and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.*, 15(16):1494–1500.
- [52] German, M. A., Luo, S., Schroth, G., Meyers, B. C., and Green, P. J. (2009). Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nat Protoc*, 4(3):356–362.
- [53] German, M. a., Pillay, M., Jeong, D.-H., Hetawal, A., Luo, S., Janardhanan, P., Kannan, V., Rymarquis, L. a., Nobuta, K., German, R., De Paoli, E., Lu, C., Schroth, G., Meyers, B. C., and Green, P. J. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature biotechnology*, 26(8):941–6.
- [54] Giegerich, R., Rehmsmeier, M., Steffen, P., and Ho, M. (2004). Fast and effective prediction of microRNA / target duplexes. *RNA*, 100(2003):1507–1517.
- [55] Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, 40(Database issue):D1178–86.

- [56] Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature*, 432(7014):235–240.
- [57] Grimson, A., Farh, K. K., Johnston, W. K., Garrett-Engele, P., Lim, L. P., and Bartel, D. P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, 27(1):91–105.
- [58] Guil, S. and Caceres, J. F. (2007). The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat. Struct. Mol. Biol.*, 14(7):591–596.
- [59] Guo, H., Ingolia, N. T., Weissman, J. S., and Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308):835–840.
- [60] Hamilton, A. J. and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441):950–952.
- [61] Han, J., Lee, Y., Yeom, K. H., Kim, Y. K., Jin, H., and Kim, V. N. (2004a). The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, 18(24):3016–3027.
- [62] Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., Sohn, S. Y., Cho, Y., Zhang, B. T., and Kim, V. N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5):887–901.
- [63] Han, M.-H., Goud, S., Song, L., and Fedoroff, N. (2004b). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(4):1093–8.
- [64] Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V. N. (2008). Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol. Cell*, 32(2):276–284.
- [65] Jha, A. and Shankar, R. (2011). Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics*, 12:636.
- [66] Jones-Rhoades, M. W. and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell*, 14(6):787 – 799.
- [67] Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annual review of plant biology*, 57:19–53.
- [68] Kakrana, A., Hammond, R., Patel, P., Nakano, M., and Meyers, B. C. (2014). sPARTA: a parallelized pipeline for integrated analysis of plant miRNA and cleaved mRNA data sets, including new miRNA target-identification software. *Nucleic Acids Res.*, 42(18):e139.
- [69] Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J. K., Staskawicz, B. J., and Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U.S.A.*, 103(47):18002–18007.
- [70] Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, 10(2):126–139.

- [71] Knauer, S., Holt, A. L., Rubio-Somoza, I., Tucker, E. J., Hinze, A., Pisch, M., Javelle, M., Timmermans, M. C., Tucker, M. R., and Laux, T. (2013). A protodermal miR394 signal defines a region of stem cell competence in the Arabidopsis shoot meristem. *Dev. Cell*, 24(2):125–132.
- [72] Kozomara, A. and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*, 42(Database issue):D68–73.
- [73] Krüger, J. and Rehmsmeier, M. (2006). Rnahybrid: microrna target prediction easy, fast and flexible. *Nucleic Acids Research*, 34(suppl 2):W451–W454.
- [74] Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.*, 19(9):1639–1645.
- [75] Kurihara, Y. and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U.S.A.*, 101(34):12753–12758.
- [76] Lai, E. C. (2004). Predicting and validating microRNA targets. *Genome Biol.*, 5(9):115.
- [77] Lanet, E., Delannoy, E., Sormani, R., Floris, M., Brodersen, P., Crete, P., Voinnet, O., and Robaglia, C. (2009). Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell*, 21(6):1762–1768.
- [78] Laufs P, Peaucelle A, M. H. and J, T. (2004). MicroRNA regulation of the CUC genes is required for boundary size control in Arabidopsis meristems. *Development*, 12(5):622–627.
- [79] Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5):843–854.
- [80] Lekprasert, P., Mayhew, M., and Ohler, U. (2011). Assessing the utility of thermodynamic features for microRNA target prediction under relaxed seed and no conservation requirements. *PLoS ONE*, 6(6):e20622.
- [81] Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, 120(1):15–20.
- [82] Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. *Cell*, 115(7):787–798.
- [83] Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes Dev.*, 17(8):991–1008.
- [84] Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002a). Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science*, 297(September):2053–2056.

- [85] Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science*, 297(5589):2053–2056.
- [86] Lobbes, D., Rallapalli, G., Schmidt, D. D., Martin, C., and Clarke, J. (2006). SERRATE: a new player on the plant microRNA scene. *EMBO reports*, 7(10):1052–8.
- [87] Lorenz, R., Bernhart, S. H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms Mol Biol*, 6:26.
- [88] Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, 303(5654):95–98.
- [89] Mallory, A. C., Reinhart, B. J., Jones-rhoades, M. W., Zamore, P. D., Kathryn, M., and Bartel, D. P. (2004). MicroRNA control of PHABULOSA in leaf development : importance of pairing to the microRNA 5' region. *Embo journal*, 23(16):3356–3364.
- [90] Manavella, P. A., Koenig, D., and Weigel, D. (2012). Plant secondary siRNA production determined by microRNA-duplex structure. *Proc. Natl. Acad. Sci. U.S.A.*, 109(7):2461–2466.
- [91] Maragkakis, M., Alexiou, P., Papadopoulos, G. L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V. A., Sethupathy, P., Vergoulis, T., Koziris, N., Sellis, T., Tsanakas, P., and Hatzigeorgiou, A. G. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, 10:295.
- [92] Maragkakis, M., Vergoulis, T., Alexiou, P., Reczko, M., Plomaritou, K., Gousis, M., Kourtis, K., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A. G. (2011). DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res.*, 39(Web Server issue):W145–148.
- [93] Markham, N. R. and Zuker, M. (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, 33(Web Server issue):W577–581.
- [94] Mateos, J. L., Bologna, N. G., Chorostecki, U., and Palatnik, J. F. (2010). Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Current Biology*, 20(1):49–54.
- [95] Maunoury, N. (2011). AGO1 and AGO2 Act Redundantly in miR408-Mediated Plantacyanin Regulation. *Plos One*, 6(12).
- [96] Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G. J., and Qi, Y. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133(1):116–127.
- [97] Michlewski, G. and Caceres, J. F. (2010). Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat. Struct. Mol. Biol.*, 17(8):1011–1018.
- [98] Millar, A. a. and Gubler, F. (2005). The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *The Plant cell*, 17(3):705–21.

- [99] Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L., Chapman, E. J., Fahlgren, N., Allen, E., and Carrington, J. C. (2008a). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 Trans -Acting siRNA Formation. *Cell*, pages 128–141.
- [100] Montgomery, T. A., Yoo, S. J., Fahlgren, N., Gilbert, S. D., Howell, M. D., Sullivan, C. M., Alexander, A., Nguyen, G., Allen, E., Ahn, J. H., and Carrington, J. C. (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. *Proc. Natl. Acad. Sci. U.S.A.*, 105(51):20055–20062.
- [101] Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443 – 453.
- [102] Niyogi, K. K., Pilon, M., Shikanai, T., Abdel-ghany, S. E., and Mu, P. (2005). Two P-Type ATPases Are Required for Copper Delivery in Arabidopsis thaliana Chloroplasts. *Plant Cell*, 17(April):1233–1251.
- [103] Notredame, C., Higgins, D. G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 302(1):205–217.
- [104] Obernosterer, G., Leuschner, P. J., Alenius, M., and Martinez, J. (2006). Post-transcriptional regulation of microRNA expression. *RNA*, 12(7):1161–1167.
- [105] Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, 130(1):89–100.
- [106] Ossowski, S., Schwab, R., and Weigel, D. (2008). Gene silencing in plants using artificial microRNAs and other small RNAs. *Plant J.*, 53(4):674–690.
- [107] Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955):257–263.
- [108] Palatnik, J. F., Wollmann, H., Schommer, C., Schwab, R., Boissbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J. C., and Weigel, D. (2007). Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. *Developmental cell*, 13(1):115–25.
- [109] Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A. G. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.*, 41(Web Server issue):W169–173.
- [110] Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408(6808):86–89.
- [111] Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, 29(9):e45.

- [112] Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & development*, 20(24):3407–25.
- [113] Ramachandran, V. and Chen, X. (2008). Small RNA metabolism in *Arabidopsis*. *Trends Plant Sci.*, 13(7):368–374.
- [114] Rampey, R. A., Leclerc, S., Kowalczyk, M., Ljung, K., Bartel, B., Biology, C., and Texas, R. A. R. (2004). A Family of Auxin-Conjugate Hydrolases That Contributes to Free Indole-3-Acetic Acid Levels during *Arabidopsis* Germination 1. *Plant Physiol.*, 135(June):978–988.
- [115] Rehmsmeier, M., Steffen, P., Hochsmann, M., and Giegerich, R. (2004). Fast and effective prediction of microRNA/target duplexes. *RNA*, 10(10):1507–1517.
- [116] Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P. (2002). MicroRNAs in plants. *Genes Dev*, pages 1616–1626.
- [117] Ren, G., Xie, M., Zhang, S., Vinovskis, C., Chen, X., and Yu, B. (2014). Methylation protects microRNAs from an AGO1-associated activity that uridylates 5' RNA fragments generated by AGO1 cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, 111(17):6365–6370.
- [118] Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell*, 110(4):513 – 520.
- [119] Rodriguez, R. E., Mecchia, M. A., Debernardi, J. M., Schommer, C., Weigel, D., and Palatnik, J. F. (2010). Control of cell proliferation in *Arabidopsis thaliana* by microRNA miR396. *Development*, 112:103–112.
- [120] Rubio-Somoza, I., Cuperus, J. T., Weigel, D., and Carrington, J. C. (2009). Regulation and functional specialization of small RNA-target nodes during plant development. *Curr. Opin. Plant Biol.*, 12(5):622–627.
- [121] Rubio-Somoza, I. and Weigel, D. (2011). MicroRNA networks and developmental plasticity in plants. *Trends Plant Sci.*, 16(5):258–264.
- [122] Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007a). Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149):83–86.
- [123] Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., and Lai, E. C. (2007b). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.*, 17(12):1850–1864.
- [124] Schapire, A. L., Bologna, N. G., Moro, B., Zhai, J., Meyers, B. C., and Palatnik, J. F. (2013). Construction of Specific Parallel Amplification of RNA Ends (SPARE) libraries for the systematic identification of plant microRNA processing intermediates. *Methods (San Diego, Calif.)*, 64(3):283–91.
- [125] Schauer, S. E., Jacobsen, S. E., Meinke, D. W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in *Arabidopsis* development. *Trends Plant Sci.*, 7(11):487–491.



- [126] Schiefthaler, Balasubramanian, Sieber, Chevalier, Wisman, and Schneitz (1999). Molecular analysis of NOZZLE, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci.*, 96(September):11664–11669.
- [127] Schwab, R., Palatnik, J. F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of micrornas on the plant transcriptome. *Developmental Cell*, 8(4):517 – 527.
- [128] Shenoy, A. and Blelloch, R. H. (2014). Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nat. Rev. Mol. Cell Biol.*, 15(9):565–576.
- [129] Song, L., Axtell, M. J., and Fedoroff, N. V. (2010). RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Curr. Biol.*, 20(1):37–41.
- [130] Song, L., Han, M. H., Lesicka, J., and Fedoroff, N. (2007). *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U.S.A.*, 104(13):5437–5442.
- [131] Souret, F. F., Kastenmayer, J. P., and Green, P. J. (2004). AtXRN4 degrades mRNA in *Arabidopsis* and its substrates include selected miRNA targets. *Mol. Cell*, 15(2):173–183.
- [132] Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T., and Pandey, S. P. (2014). A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. *BMC Genomics*, 15:348.
- [133] Stark, A., Brennecke, J., Bushati, N., Russell, R. B., and Cohen, S. M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 123(6):1133–1146.
- [134] Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of *Drosophila* MicroRNA targets. *PLoS Biol.*, 1(3):E60.
- [135] Sunkar, R., Girke, T., Jain, P. K., and Zhu, J. K. (2005). Cloning and characterization of microRNAs from rice. *Plant Cell*, 17(5):1397–1411.
- [136] Sunkar, R. and Jagadeeswaran, G. (2008). In silico identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol.*, 8:37.
- [137] Sunkar, R. and Zhu, J. K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell*, 16(8):2001–2019.
- [138] Takeda, A., Iwasaki, S., and Watanabe, T. (2008). The Mechanism Selecting the Guide Strand from Small RNA Duplexes is Different Among Argonaute Proteins. *Plant Cell Physiol*, 49(4):493–500.
- [139] Thadani, R. and Tammi, M. T. (2006). MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinformatics*, 7 Suppl 5:S20.
- [140] Todesco, M., Balasubramanian, S., Cao, J., Ott, F., Sureshkumar, S., Schneeberger, K., Meyer, R. C., Altmann, T., and Weigel, D. (2012). Natural variation in biogenesis efficiency of individual *Arabidopsis thaliana* microRNAs. *Curr. Biol.*, 22(2):166–170.

- [141] Trabucchi, M., Briata, P., Garcia-Mayoral, M., Haase, A. D., Filipowicz, W., Ramos, A., Gherzi, R., and Rosenfeld, M. G. (2009). The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature*, 459(7249):1010–1014.
- [142] Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.*, 20(7):759–771.
- [143] Vaucheret, H. (2008). Plant ARGONAUTES '. *Cell*.
- [144] Vazquez, F., Cre, P., and Bartel, D. P. (2004a). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev*, pages 1187–1197.
- [145] Vazquez, F., Gascioli, V., Cr  t  , P., and Vaucheret, H. (2004b). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Current biology : CB*, 14(4):346–51.
- [146] Viswanathan, S. R., Daley, G. Q., and Gregory, R. I. (2008). Selective blockade of microRNA processing by Lin28. *Science*, 320(5872):97–100.
- [147] Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends Genet.*, 17(8):449–459.
- [148] Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. *Cell*, 136(4):669–687.
- [149] Wang, X.-J., Reyes, J., Chua, N.-H., and Gaasterland, T. (2004). Prediction and identification of arabidopsis thaliana micrnas and their mrna targets. *Genome Biology*, 5(9):R65.
- [150] Weijers, D. and Jurgens, G. (2004). Funneling auxin action: specificity in signal transduction. *Curr. Opin. Plant Biol.*, 7(6):687–693.
- [151] Werner, S., Wollmann, H., Schneeberger, K., and Weigel, D. (2010). Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. *Curr. Biol.*, 20(1):42–48.
- [152] Wu, M. F., Tian, Q., and Reed, J. W. (2006). Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development*, 133(21):4211–4218.
- [153] Xie, F. and Zhang, B. (2010). Target-align: a tool for plant microRNA target identification. *Bioinformatics*, 26(23):3002–3003.
- [154] Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., and Carrington, J. C. (2005). Expression of Arabidopsis MIRNA Genes . *Plant Physiology*, 138(August):2145–2154.
- [155] Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation. *Current Biology*, 13(9):784–789.
- [156] Xie, Z. and Qi, X. (2008). Diverse small RNA-directed silencing pathways in plants. *Biochim. Biophys. Acta*, 1779(11):720–724.

- [157] Yamasaki, H., Abdel-Ghany, S. E., Cohu, C. M., Kobayashi, Y., Shikanai, T., and Pilon, M. (2007). Regulation of copper homeostasis by micro-RNA in Arabidopsis. *The Journal of biological chemistry*, 282(22):16369–78.
- [158] Yan, T., Yoo, D., Berardini, T. Z., Mueller, L. A., Weems, D. C., Weng, S., Cherry, J. M., and Rhee, S. Y. (2005). Patmatch: a program for finding patterns in peptide and nucleotide sequences. *Nucleic Acids Research*, 33(suppl 2):W262–W266.
- [159] Yang, C. Y., Huang, Y. H., Lin, C. P., Lin, Y. Y., Hsu, H. C., Wang, C. N., Liu, L. Y., Shen, B. N., and Lin, S. S. (2015). MicroRNA396-Targeted SHORT VEGETATIVE PHASE Is Required to Repress Flowering and Is Related to the Development of Abnormal Flower Symptoms by the Phyllody Symptoms1 Effector. *Plant Physiol.*, 168(4):1702–1716.
- [160] Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. (2006). SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J.*, 47(6):841–850.
- [161] Yang, W.-c., Ye, D., Xu, J., and Sundaresan, V. (1999). The SPOROCTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.*, pages 2108–2117.
- [162] Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.*, 17(24):3011–3016.
- [163] Yoshikawa, M., Peragine, A., Park, M. Y., and Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.*, 19(18):2164–2175.
- [164] Zeng, Y. and Cullen, B. R. (2005). Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J. Biol. Chem.*, 280(30):27595–27603.
- [165] Zhang, B., Pan, X., Cannon, C. H., Cobb, G. P., and Anderson, T. A. (2006). Conservation and divergence of plant microRNA genes. *Plant J.*, 46(2):243–259.
- [166] Zhang, L., Chia, J. M., Kumari, S., Stein, J. C., Liu, Z., Narechania, A., Maher, C. A., Guill, K., McMullen, M. D., and Ware, D. (2009). A genome-wide characterization of microRNA genes in maize. *PLoS Genet.*, 5(11):e1000716.
- [167] Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., Zhang, X., and Jin, H. (2010). Multiple distinct small RNAs originate from the same microRNA precursors. *Genome Biol.*, 11(8):R81.
- [168] Zheng, Y., Li, Y. F., Sunkar, R., and Zhang, W. (2012). SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Res.*, 40(4):e28.
- [169] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13):3406–3415.

# Anexo

## Código fuente 9.1: circos.conf

---

```
<<include etc/colors_fonts_patterns.conf>>
<<include etc/housekeeping.conf>>

<image>
    angle_offset*    = 114.5
    <<include etc/image.conf>>
</image>

# Chromosome name, size and color definition
karyotype = data/karyotype.txt

<ideogram>
    <spacing>
        default = 0.001r
        <pairwise last first>
            spacing = 200r
        </pairwise>
    </spacing>

    label_format      = eval( if(var(label) =~ /(\\d*)(\\w*)/){ if($2 eq 'x'){ sprintf "%d —", $1;} else{ sprintf "%d %s", $1, $2;}})
    radius             = 0.80r
    thickness          = 35p
    fill               = yes
    stroke_color       = dgrey
    stroke_thickness   = 0p
    show_label         = yes
    label_font         = semibold
    label_radius       = 1r + 15p
    label_color        = grey
    label_size         = 20
    label_parallel     = no

</ideogram>

<<include data/links.conf>>

<highlights>
    <highlight>
        file           = data/highlight.txt
        fill_color     = 96,122,66
        r0 = 1.10r
        r1 = 1.12r
    </highlight>
</highlights>

<plots>
    type              = text
    color             = black
    label_font        = default
```

```

label_size = 40p

<plot>

    file = data/text_5p_3p.txt
    r0 = 1r+80p
    r1 = 1r+200p
    label_parallel = yes

    show_links = yes
    link_dims = 4p,4p,8p,4p,4p
    link_thickness = 2p
    link_color = black

</plot>

<<include data/histogram.conf>>

</plots>

```

---

### Código fuente 9.2: histogram.conf

---

```

<plot>

    type = histogram
    file = data/histogram.txt
    r1 = 0.97r
    r0 = 0.8r
    max = 0.5
    min = -0.5
    stroke_type = outline
    thickness = 0
    color = vdgrey
    extend_bin = no

    <backgrounds>
        <background>
            color = 231,231,231
        </background>
    </backgrounds>

    <axes>
        <axis>
            spacing = 0.2r
            color = lgrey
            thickness = 1
        </axis>
    </axes>

    <rules>
        <rule>
            condition = 1
            fill_color = eval(sprintf("circos_hist-14-div-%d", remap_int(var(value), -0.5, 0.5, 1, 14)))
        </rule>
    </rules>

</plot>

```

---

### Código fuente 9.3: links.conf

---

```

<links>
    z = 10
    radius = 0.79r

    <link>
        bezier_radius = undef

```

```
show      = yes
color     = 197,197,197
file      = link.txt

<rules>
  <rule>
    condition = var(thickness) == 8
    thickness = 3
    color     = circos_hist-14-div-14
    z         = 15
  </rule>

  <rule>
    condition = var(thickness) <= 4
    show = no
  </rule>

  <rule>
    condition = 1
    thickness = 3
    z         = 10
  </rule>
</rules>

</link>

</links>
```

---

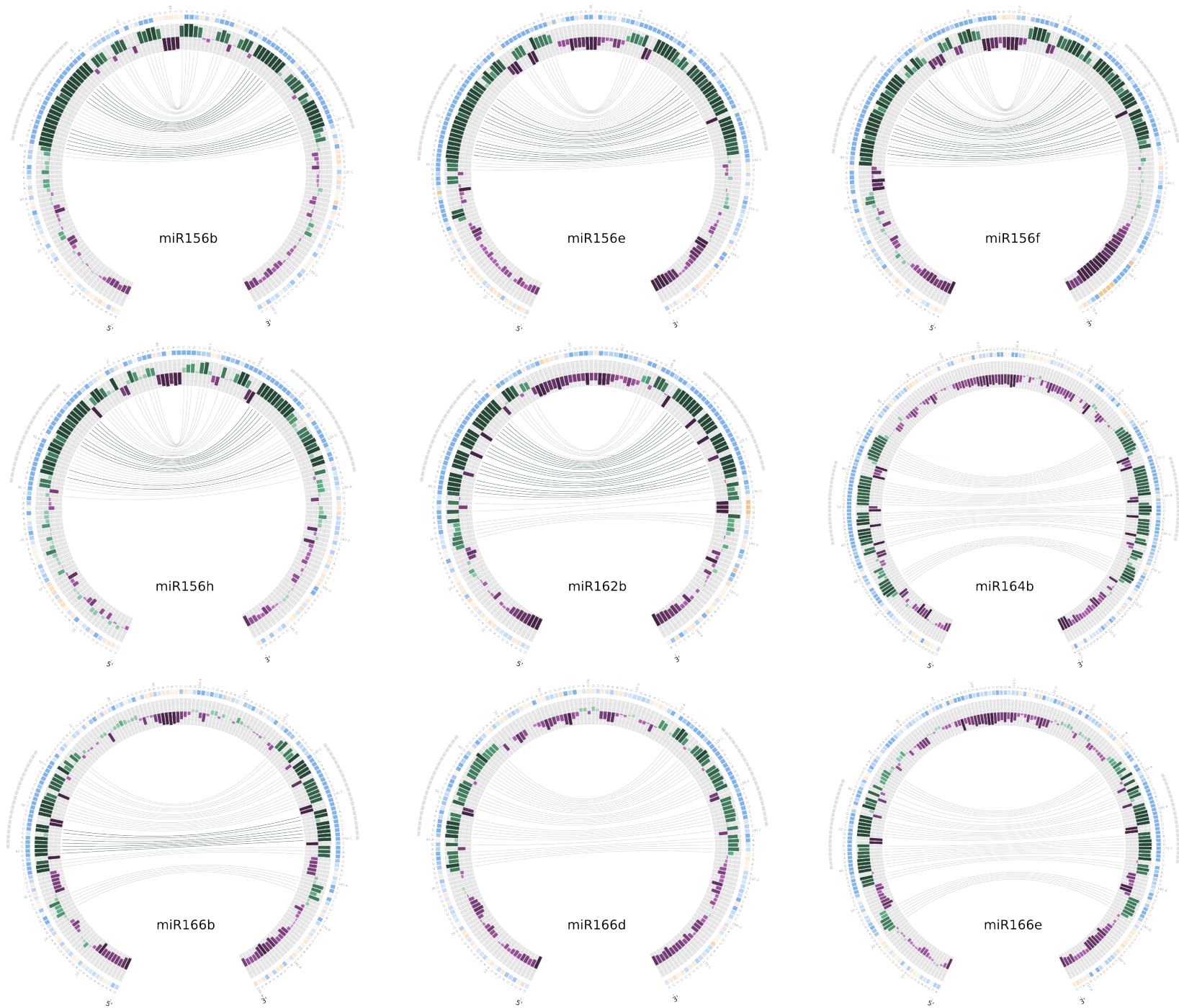


Figura 9.1: Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas. Fig 1/5.

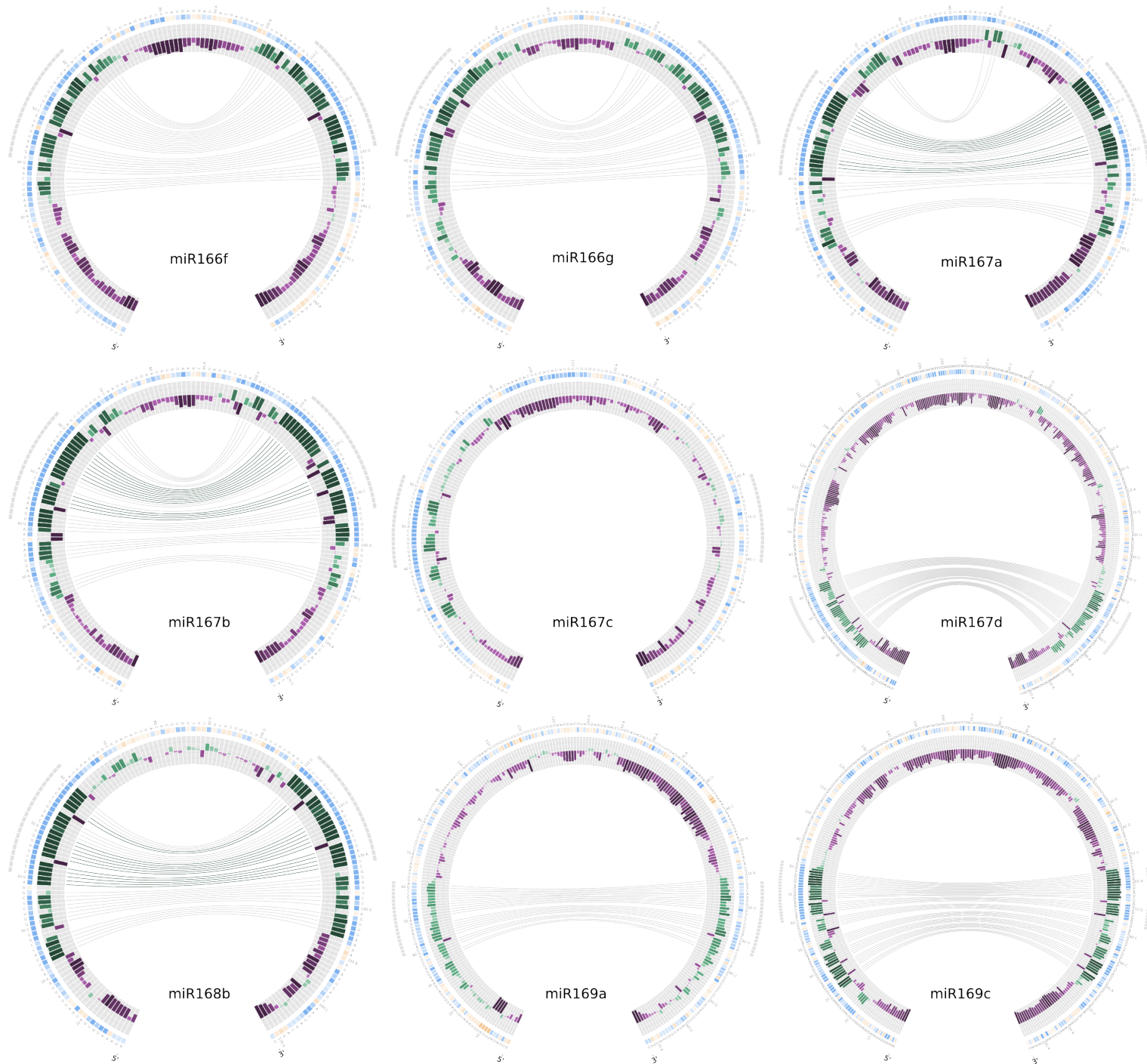


Figura 9.2: Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas. Fig 2/5.



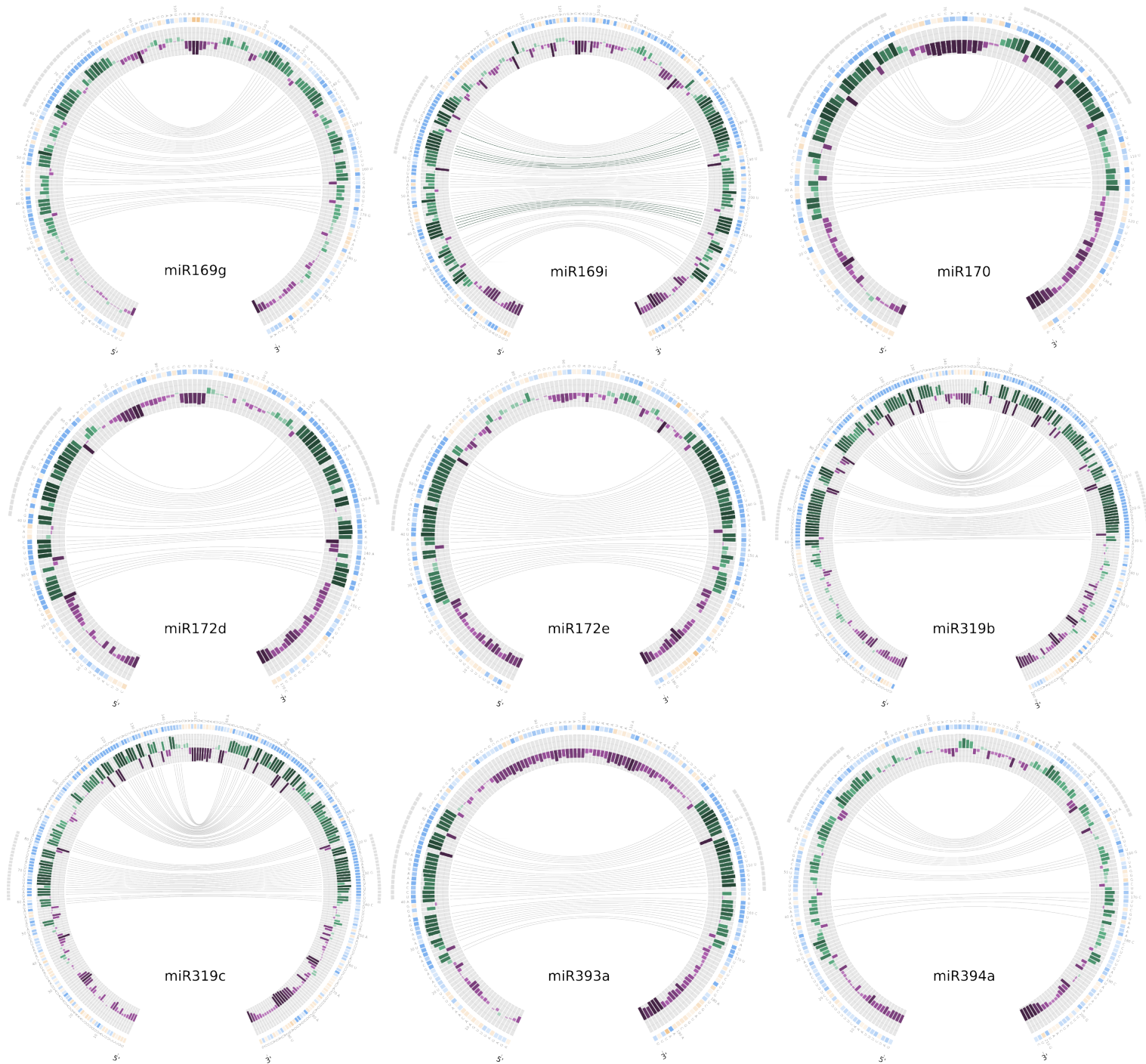


Figura 9.3: Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas. Fig 3/5.

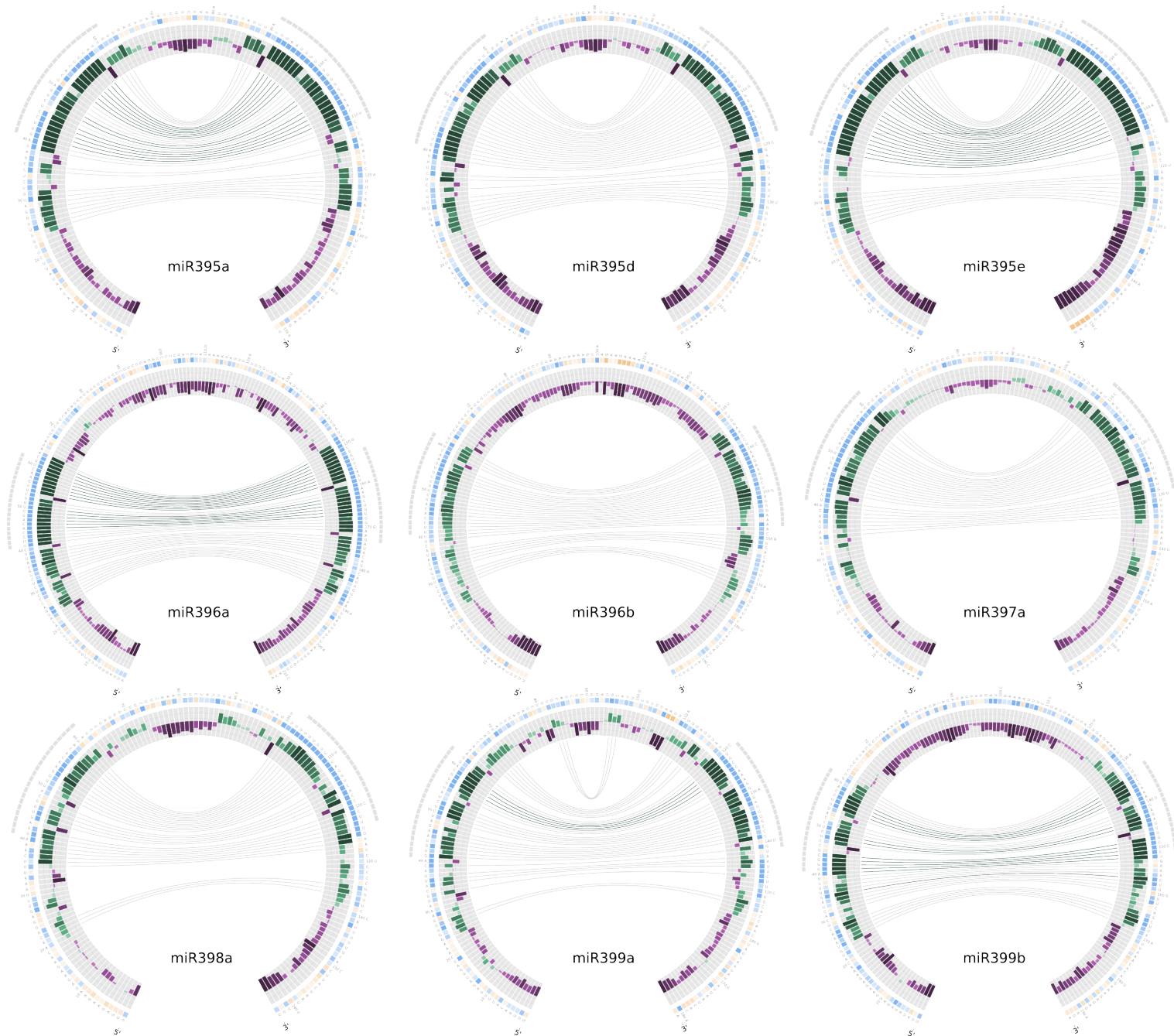


Figura 9.4: Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas. Fig 4/5.

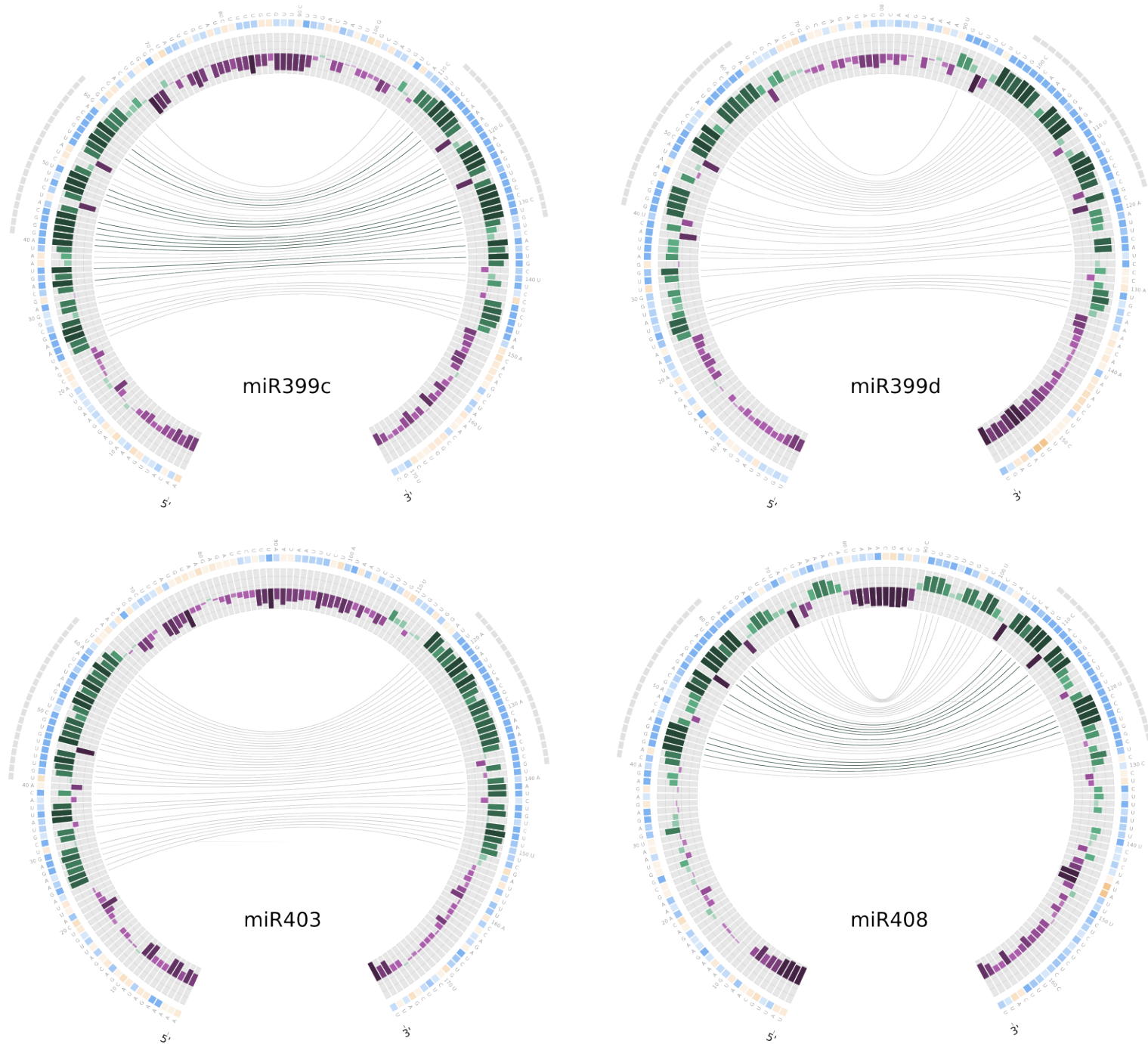


Figura 9.5: Circos de dicotiledóneas donde se detectan ortólogos en más de 20 especies de las 30 analizadas. Fig 5/5.

Tabla 9.1: Genes blanco validados de familias miARNs conservados en plantas

microRNA	Target	ID
miR156/miR157	SPL	At1g27370
miR156/miR157	SPL	At1g53160
miR156/miR157	SPL	At2g33810
miR156/miR157	SPL	At3g15270
miR156/miR157	SPL	At5g43270
miR156/miR157	SPL	At1g69170
miR156/miR157	SPL	At2g42200
miR156/miR157	SPL	At3g57920
miR156/miR157	SPL	At5g50670
miR159/miR319	TCP	At1g30210
miR159/miR319	TCP	At1g53230
miR159/miR319	TCP	At2g31070
miR159/miR319	MYB	At3g11440
miR159/miR319	TCP	At3g15030
miR159/miR319	TCP	At4g18390
miR159/miR319	MYB	At5g06100
miR159/miR319	MYB	At2g26950
miR159/miR319	MYB	At2g32460
miR159/miR319	MYB	At5g55020
miR160	ARF	At1g77850
miR160	ARF	At2g28350
miR160	ARF	At4g30080
miR162	DCL	At1g01040
miR164	NAC	At1g56010
miR164	NAC	At3g15170
miR164	NAC	At5g07680
miR164	NAC	At5g53950
miR164	NAC	At5g61430
miR164	NAC	At3g12977
miR164	NAC	At5g39610
miR166/miR165	HD-ZIPIII	At1g30490
miR166/miR165	HD-ZIPIII	At1g52150
miR166/miR165	HD-ZIPIII	At2g34710
miR166/miR165	HD-ZIPIII	At5g60690
miR166/miR165	HD-ZIPIII	At4g32880
miR167	ARF	At1g30330
miR167	ARF	At5g37020
miR168	AGO	At1g48410
miR169	HAP2	At1g17590
miR169	HAP2	At1g54160
miR169	HAP2	At1g72830
miR169	HAP2	At3g05690
miR169	HAP2	At3g20910
miR169	HAP2	At5g06510
miR170/miR171	SCL	At2g45160
miR170/miR171	SCL	At3g60630
miR170/miR171	SCL	At4g00150
miR172	AP2	At2g28550
miR172	AP2	At4g36920
miR172	AP2	At5g60120
miR172	AP2	At5g67180
miR172	AP2	At2g39250
miR172	AP2	At3g54990
miR390/miR391	TAS3	At3g17185
miR390/miR391	TAS3	At5g49615
miR390/miR391	TAS3	At5g57735
miR393	TIR1/AFB	At1g12820
miR393	bHLH	At3g23690
miR393	TIR1/AFB	At3g26810
miR393	TIR1/AFB	At3g62980
miR393	TIR1/AFB	At4g03190
miR394	F-Box	At1g27340
miR395	APS	At3g22890
miR395	AST	At5g10180
miR395	APS	At5g43780
miR395	APS	At4g14680
miR396	GRF	At2g22840
miR396	GRF	At2g36400
miR396	GRF	At2g45480
miR396	GRF	At4g24150
miR396	GRF	At4g37740
miR396	GRF	At5g53660
miR396	GRF	At3g52910
miR397	LAC	At2g29130
miR397	LAC	At2g38080
miR397	LAC	At5g60020
miR398	CSD	At1g08830
miR398	CSD	At2g28190
miR398	CytC oxidase	At3g15640
miR399	E2-UBC	At2g33770
miR399	E2-UBC	At2g33770
miR408	LAC	At2g30210
miR408	PLC	At2g02850
miR827	SPX	At1g02860

Tabla 9.2: Mecanismos de procesamiento de miARNs conservados en *A. thaliana*.

miARN	Mecanismo
miR156a	Corto de loop a base
miR156b	Corto de loop a base
miR156c	Corto de loop a base
miR156d	Corto de loop a base
miR156e	Corto de loop a base
miR156h	Corto de loop a base
miR157a	Corto de loop a base
miR157b	Corto de loop a base
miR157c	Corto de loop a base
miR159a	Secuencial de loop a base
miR159b	Secuencial de loop a base
miR160a	Corto de loop a base
miR160b	Corto de loop a base
miR160c	Corto de loop a base
miR162a	Corto de loop a base
miR164b	Corto de base a loop
miR164c	Corto de base a loop
miR165a	Corto de base a loop
miR165b	Corto de base a loop
miR166a	Corto de base a loop
miR166b	Corto de base a loop
miR166f	Corto de base a loop
miR167a	Corto de base a loop
miR167b	Corto de base a loop
miR167c	Corto de base a loop
miR167d	Corto de base a loop
miR168a	Corto de base a loop
miR168b	Corto de base a loop
miR169a	Corto de base a loop
miR169b	Secuencial de base a loop
miR169d	Secuencial de base a loop
miR169e	Secuencial de base a loop
miR169f	Secuencial de base a loop
miR169h	Secuencial de base a loop
miR169i	Secuencial de base a loop
miR169j	Secuencial de base a loop
miR169k	Secuencial de base a loop
miR169l	Secuencial de base a loop
miR169m	Secuencial de base a loop
miR169n	Secuencial de base a loop
miR170	Corto de base a loop
miR171a	Corto de base a loop
miR171b	Corto de loop a base
miR171c	Corto de loop a base
miR172a	Corto de base a loop
miR172b	Corto de base a loop
miR172c	Corto de base a loop
miR172d	Corto de base a loop
miR172e	Corto de base a loop
miR319a	Secuencial de loop a base
miR319b	Secuencial de loop a base
miR319c	Secuencial de loop a base
miR390a	Corto de base a loop
miR390b	Corto de base a loop
miR391	Corto de base a loop
miR393a	Corto de base a loop
miR393b	Corto de base a loop
miR394a	Secuencial de base a loop
miR394b	Secuencial de base a loop
miR395a	Corto de base a loop
miR395b	Corto de base a loop
miR395c	Corto de base a loop
miR395d	Corto de base a loop
miR395f	Corto de base a loop
miR396a	Corto de base a loop
miR396b	Corto de base a loop
miR397a	Corto de base a loop
miR397b	Corto de base a loop
miR397c	Corto de base a loop
miR398b	Corto de base a loop
miR398c	Corto de base a loop
miR399a	Corto de base a loop
miR399b	Corto de base a loop
miR399c	Corto de base a loop
miR399d	Corto de base a loop
miR408	Corto de base a loop
miR827	Corto de base a loop